

20th Edition

HARRISON'S

PRINCIPLES OF
INTERNAL
MEDICINE

JAMESON
FAUCI
KASPER
HAUSER
LONGO
LOSCALZO

VOLUME 1

Mc
Graw
Hill
Education

HARRISON'S
INTERNAL MEDICINE

20th Edition

VOLUME
1

Chapters 1-223

JAMESON
FAUCI
KASPER
HAUSER
LONGO
LOSCALZO

Mc
Graw
Hill
Education

HARRISON'S
INTERNAL MEDICINE

20th Edition

VOLUME
2

Chapters 224-526

JAMESON
FAUCI
KASPER
HAUSER
LONGO
LOSCALZO

Mc
Graw
Hill
Education

Chapter 1: The Practice of Medicine

The Editors

FIGURE 1-1

ENDURING VALUES OF THE MEDICAL PROFESSION

No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding.... Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.

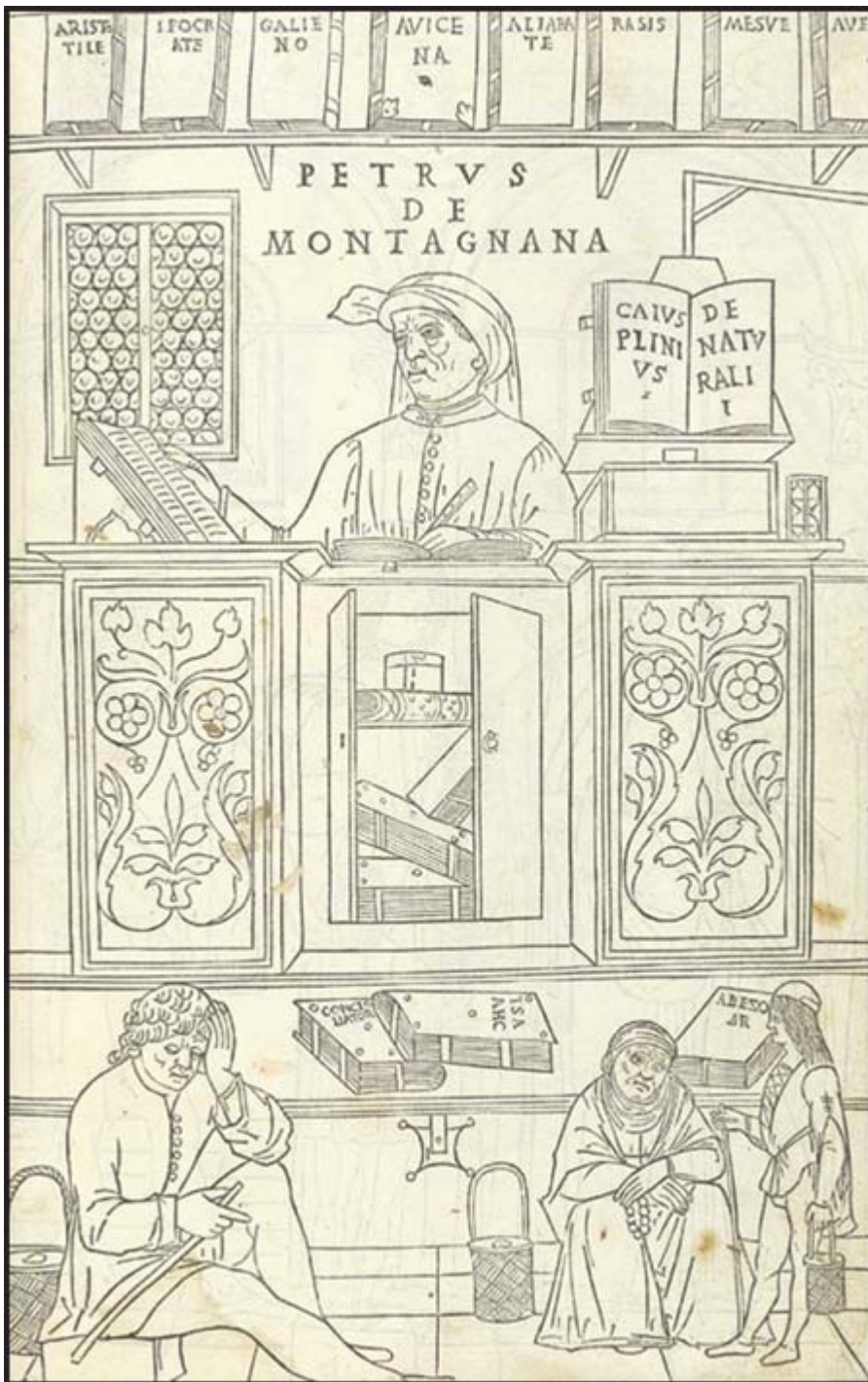
—Harrison's Principles of Internal Medicine, 1950

The practice of medicine has changed in significant ways since the first edition of this book appeared in 1950. The advent of molecular genetics, sophisticated new imaging techniques, robotics, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has changed fundamentally the way physicians define, diagnose, treat, and attempt to prevent disease. This growth of scientific knowledge is ongoing and accelerating.

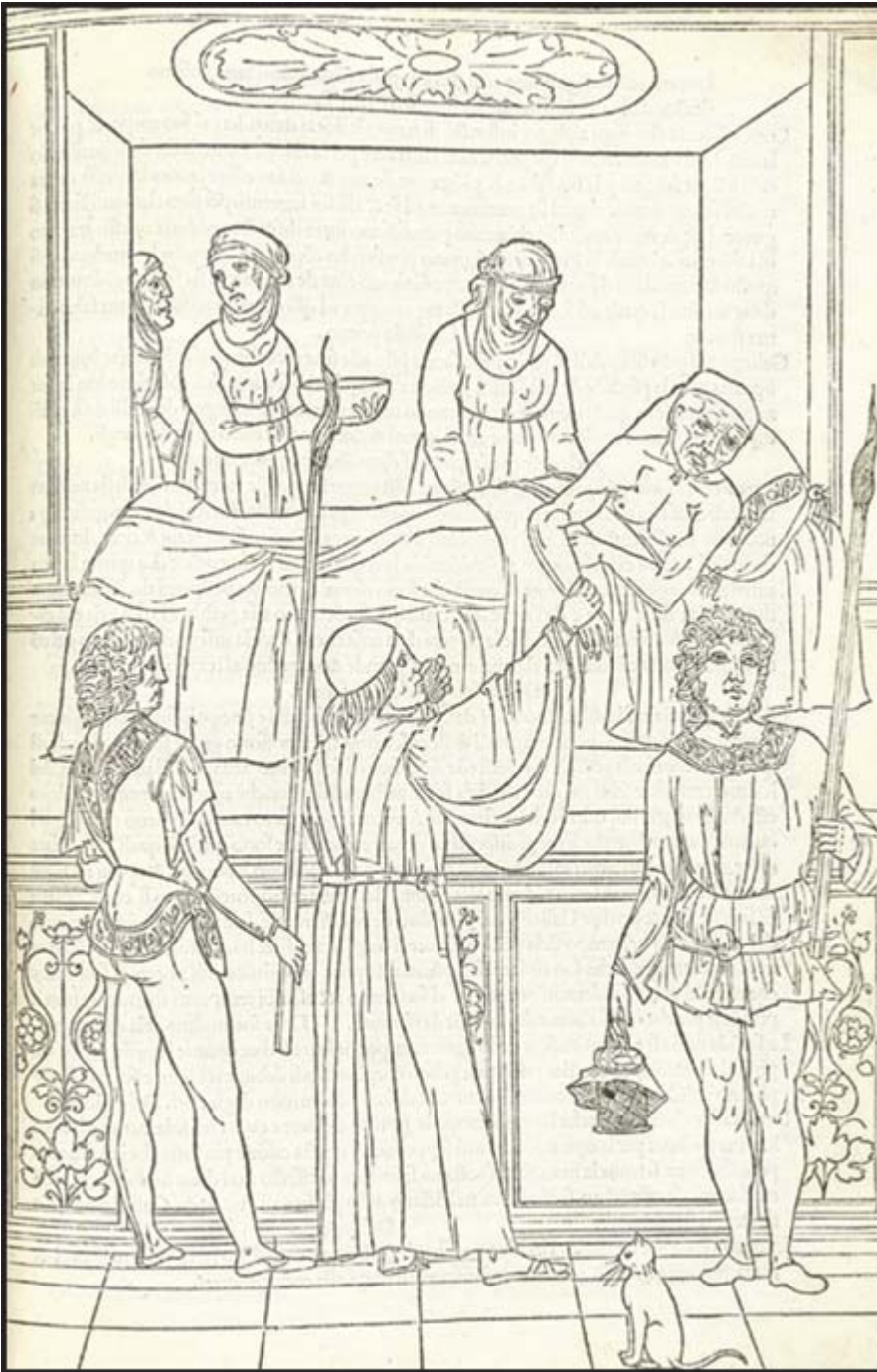
The widespread use of electronic medical records and the Internet have altered the way physicians access and exchange information as a routine part of medical practice (**Fig. 1-1**). As today's physicians strive to integrate copious amounts of scientific knowledge into everyday practice, it is critically important to remember two things: first, the ultimate goal of medicine is to prevent disease and, when it occurs, to diagnose it early and provide effective treatment; and second, despite nearly 70 years of scientific advances since the first edition of this text, a trusting relationship between physician and patient still lies at the heart of successful patient care.

FIGURE 1-1

Woodcuts from Johannes de Ketham's *Fasciculus Medicinae*, the first illustrated medical text ever printed, show methods of information access and exchange in medical practice during the early Renaissance. Initially published in 1491 for use by medical students and practitioners, *Fasciculus Medicinae* appeared in six editions over the next 25 years. *Left*: Petrus de Montagnana, a well-known physician and teacher at the University of Padua and author of an anthology of instructive case studies, consults medical texts dating from antiquity up to the early Renaissance. *Right*: A patient with plague is attended by a physician and his attendants. (Courtesy, U.S. National Library of Medicine.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine. 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

THE SCIENCE AND ART OF MEDICINE

Deductive reasoning and applied technology form the foundation for the solution to many clinical problems. Spectacular advances in biochemistry, cell biology, and genomics, coupled with newly developed imaging techniques, allow access to the innermost parts of the cell and provide a window into the most remote recesses of the body. Revelations about the nature of genes and single cells have opened a portal for formulating a new molecular basis for the physiology of systems. Increasingly, physicians are learning how subtle changes in many different genes can affect the function of cells and organisms. Researchers are

deciphering the complex mechanisms by which genes are regulated. Clinicians have developed a new appreciation of the role of stem cells in normal tissue function, in the development of cancer and other disorders, and in the treatment of certain diseases. Entirely new areas of research, including studies of chronobiology, the human microbiome, and epigenetics, have become important for understanding both health and disease. Information technology enables the interrogation of medical records from millions of individuals, yielding new insights into the etiology, characteristics, and stratification of many diseases. The knowledge gleaned from the ***science of medicine*** continues to enhance the understanding by physicians of complex pathologic processes and to provide new approaches to disease prevention, diagnosis, and treatment. Yet skill in the most sophisticated applications of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician.

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from densely populated computer screens to determine whether to treat or to “watch.” As the number of tests increases, so does the likelihood that some incidental finding, completely unrelated to the clinical problem at hand, will be uncovered. Deciding whether a clinical clue is worth pursuing or should be dismissed as a “red herring” and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the ***art of medicine***, which is as necessary to the practice of medicine as is a sound scientific base.

CLINICAL SKILLS

History-Taking

The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. Patients should, at some early point, have the opportunity to tell their own story of the illness without frequent interruption and, when appropriate, should receive expressions of interest, encouragement, and empathy from the physician. Any event related by a patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. A methodical review of systems is important to elicit features of an underlying disease that might not be mentioned in the patient’s narrative. In general, patients who feel comfortable with the physician will offer more complete information; thus, putting the patient at ease contributes substantially to obtaining an adequate history.

An informative history is more than an orderly listing of symptoms. By listening to patients and noting the way in which they describe their symptoms, physicians can gain valuable insight. Inflections of voice, facial expression, gestures, and attitude (i.e., “body language”) may offer important clues to patients’ perception of their symptoms. Because patients vary considerably in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered and can identify practical considerations for subsequent management. The family history not only identifies rare Mendelian disorders but often reveals risk factors for common disorders, such as coronary heart disease, hypertension,

autoimmunity, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy. An experienced clinician can usually formulate a relevant differential diagnosis from the history alone, using the physical examination and diagnostic tests to narrow the list or reveal unexpected findings that lead to more focused inquiry.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance a unique bond that forms the basis for a good patient–physician relationship. This process helps the physician develop an appreciation of the patient’s view of the illness, the patient’s expectations of the physician and the health care system, and the financial and social implications of the illness for the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the encounter. A hurried approach may lead patients to believe that what they are relating is not of importance to the physician, and thus they may withhold relevant information. The confidentiality of the patient–physician relationship cannot be overemphasized.

Physical Examination

The purpose of the physical examination is to identify physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient’s history. At times, however, physical signs may be the only evidence of disease and may not have been suggested by the history.

The physical examination should be methodical and thorough, with consideration given to the patient’s comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. The results of the examination, like the details of the history, should be recorded at the time they are elicited—not hours later, when they are subject to the distortions of memory. Physical examination skills should be learned under direct observation of experienced clinicians. Even highly experienced clinicians can benefit from ongoing coaching and feedback. Simulation laboratories and standardized patients play an increasingly important role in the development of clinical skills. Although the skills of physical diagnosis are acquired with experience, it is not merely technique that determines success in identifying signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers, but of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants.

Given the many highly sensitive diagnostic tests now available (particularly imaging techniques), it may be tempting to place less emphasis on the physical examination. Indeed, many patients are seen by consultants after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since important clinical findings may have escaped detection. The act of examining (touching) the patient also offers an opportunity for communication and may have reassuring effects that foster the patient–physician relationship.

Diagnostic Studies

Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems. However, accumulated results do not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their apparent precision, these tests often gain an aura of certainty regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting the tests. Physicians must weigh the expense involved in laboratory procedures against the value of the information these procedures are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request “batteries” of multiple tests, which often prove useful and can be performed with a single specimen at relatively low cost. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting a diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism or an underlying malignancy.

The thoughtful use of screening tests (e.g., measurement of low-density lipoprotein cholesterol) may allow early intervention to prevent disease ([Chap. 4](#)). Screening tests are most informative when they are directed toward common diseases and when their results indicate whether other useful—but often costly—tests or interventions are needed. On the one hand, biochemical measurements, together with simple laboratory determinations such as routine serum chemistries, blood counts, and urinalysis, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional screening-test abnormalities that do not necessarily connote significant disease. An in-depth workup after the report of an isolated laboratory abnormality in a person who is otherwise well is often wasteful and unproductive. Because so many tests are performed routinely for screening purposes, it is not unusual for one or two values to be slightly abnormal. Nevertheless, even if there is no reason to suspect an underlying illness, tests yielding abnormal results ordinarily are repeated to rule out laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient’s condition and other test results.

There is almost continual development of technically improved imaging studies with greater sensitivity and specificity. These tests provide remarkably detailed anatomic information that can be pivotal in informing medical decision-making. Ultrasonography, CT, MRI, a variety of isotopic scans, and positron emission tomography (PET) have supplanted older, more invasive approaches and opened new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies revealed findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may be associated with risk and/or expense and may lead to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

Team-Based Care

Medical practice has long involved teams, particularly physicians working with nurses. Advances in medicine have increased our ability to manage very complex clinical situations (e.g., intensive care units [ICUs], bone marrow transplantation) and have shifted the burden of disease toward chronic illnesses. Because an individual patient may have multiple chronic diseases, he or she may be cared for by different specialists as well as a primary care physician. In the inpatient setting, care may involve multiple consultants along with the primary admitting physician. Communication through the medical record is necessary but not sufficient, particularly when patients have complex medical problems or when difficult decisions need to be made about the optimal management plan. Physicians should willingly meet face-to-face or by phone to ensure clear communication and thoughtful planning. It is important to note that patients often receive or perceive different messages from various care providers; attempts should be made to provide consistency among these messages to the patient. Management plans and treatment options should be outlined succinctly and clearly for the patient.

Another dimension of team-based care involves allied health professions. It is not unusual for a hospitalized patient to encounter physical therapists, pharmacists, respiratory therapists, radiology technicians, social workers, dietitians, and transport personnel (among others) in addition to physicians and nurses. Each of these individuals contributes to clinical care as well as to the patient's experience with the health care system. In the outpatient setting, disease screening and chronic disease management are often carried out by nurses, physician assistants, or other allied health professionals.

The growth of team-based care has important implications for medical culture, student and resident training, and the organization of health care systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most challenging aspect of implementing team-based care. While communication can be aided by electronic devices, including medical records, apps, or text messages, it is vitally important to balance efficiency with taking the necessary time to speak directly with colleagues.

The Dichotomy of Inpatient and Outpatient Internal Medicine

The hospital environment has experienced sweeping changes over the last few decades. Emergency departments and critical care units have evolved to manage critically ill patients, allowing them to survive formerly fatal conditions. In parallel, there is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by reducing inpatient hospitalization.

In addition to traditional medical beds, hospitals now encompass multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units. A consequence of this differentiation has been the emergence of new specialties (e.g., emergency medicine and end-of-life care) and the provision of in-hospital care by hospitalists and intensivists. Most *hospitalists* are board-certified internists who bear primary responsibility for the care of hospitalized patients

and whose work is limited entirely to the hospital setting. The shortened length of hospital stay means that most patients receive only acute care while hospitalized; the increased complexities of inpatient medicine make the presence of an internist with specific training, skills, and experience in the hospital environment extremely beneficial. *Intensivists* are board-certified physicians who are further certified in critical care medicine and who direct and provide care for very ill patients in critical care units. Clearly, an important challenge in internal medicine today is to ensure the continuity of communication and information flow between a patient's primary care physician and those who are in charge of the patient's hospital care. Maintaining these channels of communication is frequently complicated by patient "handoffs"—i.e., transitions from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, from a medical to a surgical service and vice versa, and from the hospital to the outpatient environment.

The involvement of many care providers in conjunction with these transitions can threaten the traditional one-to-one relationship between patient and primary care physician. Of course, patients can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to provide cohesive guidance through an illness*. To meet this challenge, primary care physicians must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals who care for their patients in the hospital. In addition, primary care physicians must ensure that their patients benefit from scientific advances and the expertise of specialists, both in and out of the hospital. Primary care physicians should explain the role of these specialists to reassure patients that they are in the hands of physicians best trained to manage an acute illness. However, the primary care physician should assure patients and their families that decisions are being made in consultation with these specialists. The evolving concept of the "medical home" incorporates team-based primary care with subspecialty care in a cohesive environment that ensures smooth transitions of care.

Mitigating the Stress of Acute Illness

Few people are prepared for a new diagnosis of cancer or anticipate the occurrence of a myocardial infarction, stroke, or major accident. The care of a frightened or distraught patient is confounded by these understandable responses to life-threatening events. The physician and other health providers can reduce the shock of life-changing events by providing information in a clear, calm, consistent, and reassuring manner. Often, information and reassurance need to be repeated. Caregivers should also recognize that, for outsiders, hospital emergency rooms, operating rooms, ICUs, and general medical floors represent an intimidating environment. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—hospitalists, specialists, nurses, nurses' aides, physicians' assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended at times; and they may be obligated to share a room with other patients who have their own health problems. It is little wonder that patients may be stressed by this environment. Physicians who appreciate the hospital experience from the patient's perspective and who

make an effort to guide the patient through this experience may make a stressful situation more tolerable and enhance the patient's chances for an optimal recovery.

Medical Decision-Making

Medical decision-making is a fundamental responsibility of the physician and occurs at each stage of the diagnostic and therapeutic process. The decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases for a given patient. Application of the scientific method, including hypothesis formulation and data collection, is essential to the process of accepting or rejecting a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Whenever possible, decisions should be evidence-based, taking advantage of rigorously designed clinical trials or objective comparisons of different diagnostic tests. *Evidence-based medicine* is in sharp contrast to anecdotal experience, which is often biased. Unless attuned to the importance of using larger, objective studies for making decisions, even the most experienced physicians can be influenced to an undue extent by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of routine medical practice and has led to the publication of many useful practice guidelines.

Despite the importance of evidence-based medicine, much medical decision-making still relies on good clinical judgment, an attribute that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors, along with the inevitable uncertainties, and then making a sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable in synthesizing the available information, including diagnostic tests, Bayes' theorem, and multivariate statistical models. Diagnostic tests serve to reduce uncertainty about an individual's diagnosis or prognosis and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and the physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate) as well as the predictive value of a positive and a negative result. **See Chap. 3 for a more thorough discussion of decision-making in clinical medicine.**

Practice Guidelines

Many professional organizations and government agencies have developed formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. Clinical guidelines can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. These guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-

practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with different perspectives may develop divergent recommendations regarding issues as basic as the need for screening of women by mammography or of men by serum prostate-specific antigen (PSA). Finally, guidelines, as the term implies, do not—and cannot be expected to—account for the uniqueness of each individual and his or her illness. The physician’s challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Precision Medicine

The concept of *precision* or *personalized medicine* reflects the growing recognition that diseases once lumped together can be further stratified on the basis of genetic, biomarker, phenotypic, and/or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations. Inherent in this concept is the goal of targeting therapies in a more specific way to improve clinical outcomes for the individual patient and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the evolution of clinical practice guidelines, which are usually developed for populations of patients or a particular diagnosis (e.g., hypertension, thyroid nodule). As the pathophysiology, prognosis, and treatment responses of subgroups within these diagnoses become better understood, the relevant clinical guidelines incorporate progressively more refined recommendations for individuals within these subgroups. The role of precision medicine is particularly important for cancers in which genetic testing is able to predict responses (or the lack thereof) to targeted therapies ([Chap. 69](#)). One can anticipate similar applications of precision medicine in pharmacogenomics, immunologic disorders, and diseases in which biomarkers better predict treatment responses.

Evaluation of Outcomes

Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. These measures may oversimplify the complexity of a clinical condition as patients often present with a major clinical problem in the context of multiple complicating background illnesses. For example, a patient may present with chest pain and cardiac ischemia, but with a background of chronic obstructive pulmonary disease and renal insufficiency. For this reason, outcome measures such as mortality, length of hospital stay, or readmission rates are typically risk-adjusted. An important point to remember is that patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient’s health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important domains can be assessed through structured interviews or specially designed questionnaires. Such assessments provide useful parameters by which a physician can judge patients’ subjective views of their disabilities and responses to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Many health systems use survey and patient feedback data to assess qualitative features such as patient satisfaction, access to care, and communication with nurses and physicians. In the United States, HCAHPS

(Hospital Consumer Assessment of Healthcare Providers and Systems) surveys are used by many systems and are publically reported. Social media is also being used to assess feedback in real time as well as to share patient experiences with health care systems.

Errors in the Delivery of Health Care

A series of reports from the Institute of Medicine (now the National Academy of Medicine [NAM]) called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems. It is the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that rely on electronic processes or, when electronic options are not available, that eliminate misreading of handwriting. Whatever the clinical situation, it is the physician's responsibility to use powerful therapeutic measures wisely, with due regard for their beneficial actions, potential dangers, and cost. Implementation of infection control systems, enforcement of hand-washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections. Central-line infection rates have been dramatically reduced at many centers by careful adherence of trained personnel to standardized protocols for introducing and maintaining central lines. Rates of surgical infection and wrong-site surgery can likewise be reduced by the use of standardized protocols and checklists. Falls by patients can be minimized by judicious use of sedatives and appropriate assistance with bed-to-chair and bed-to-bathroom transitions. Taken together, these and other measures are saving thousands of lives each year.

Electronic Medical Records

Both the growing reliance on computers and the strength of information technology now play central roles in medicine, including efforts to reduce medical errors. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records (EMRs), computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive, sending reminders or warning of potential medical errors.

EMRs offer rapid access to information that is invaluable in enhancing health care quality and patient safety, including relevant data, historical and clinical information, imaging studies, laboratory results, and medication records. These data can be used to monitor and reduce unnecessary variations in care and to provide real-time information about processes of care and clinical outcomes. Ideally, patient records are easily transferred across the health care system. However, technological limitations and concerns about privacy and cost continue to limit broad-based use of EMRs in many clinical settings.

For all of the advantages of EMRs, they can create distance between the physician and patient if care is not taken to preserve face-to-face contact. EMRs also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as "cutting and pasting" parts of earlier notes into the daily record, thereby increasing the risk of errors. EMRs also structure information in a manner that disrupts the traditional narrative flow across time and among providers. These features, which may be frustrating for some providers, must be weighed against the advantages of ready access to past medical history, imaging, laboratory data, and consultant notes.

It is important to emphasize that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. Clinical knowledge and an understanding of a patient's needs, supplemented by quantitative tools, still represent the best approach to decision-making in the practice of medicine.

THE PATIENT–PHYSICIAN RELATIONSHIP

*The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity, **for the secret of the care of the patient is in caring for the patient.***

—Francis W. Peabody, October 21, 1925, Lecture at Harvard Medical School

Physicians must never forget that patients are individuals with problems that all too often transcend their physical complaints. They are not “cases” or “admissions” or “diseases.” Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and offer reassurance, but they must never come across as arrogant or patronizing. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. The physician needs to consider the setting in which an illness occurs—in terms not only of patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient–physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

Informed Consent

The fundamental principles of medical ethics require physicians to act in the patient's best interest and to respect the patient's autonomy. These requirements are particularly relevant to the issue of informed consent. Patients are required to sign consent forms for most diagnostic or therapeutic procedures. Many patients possess limited medical knowledge and must rely on their physicians for advice. Communicating in a clear and understandable manner, physicians must fully discuss the alternatives for care and explain the risks, benefits, and likely consequences of each alternative. The physician is responsible for ensuring that the patient thoroughly understands these risks and benefits; encouraging questions is an important part of this process. It may be necessary to go over certain issues with the patient more than once. This is the very definition of *informed consent*. Complete, clear explanation and discussion of the proposed procedures and treatment can greatly mitigate the fear of the unknown that commonly accompanies hospitalization. Often the patient's understanding is enhanced by repeatedly discussing the issues in an unthreatening and supportive way, answering new questions that occur to the patient as they arise. Clear communication can also help alleviate misunderstandings in situations where complications of intervention occur.

Special care should be taken to ensure that a physician seeking a patient's informed consent has no real or apparent conflict of interest.

Approach to Grave Prognoses and Death

No circumstance is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to optimize quality of life?

Transparency of information, delivered in an appropriate manner, is essential in the face of a terminal illness. Even patients who seem unaware of their medical circumstances, or whose family members have protected them from diagnoses or prognoses, often have keen insights into their condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to talk with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know and when he or she wants to know it. On the basis of the patient's responses, the physician can assess the right tempo for sharing information. Ultimately, the patient must understand the expected course of the disease so that appropriate plans and preparations can be made. The patient should participate in decision-making with an understanding of the goal of treatment (palliation) and its likely effects. The patient's religious beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

The physician should provide or arrange for emotional, physical, and spiritual support and must be compassionate, unhurried, and open. In many instances, there is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining equipment can detract from attention to the whole person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first hasten to help, is a guiding principle. In offering care to a dying patient, a physician should be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt or even anger. It is important for the physician to assure the family that everything reasonable is being done. A substantial challenge in these discussions is that the physician often does not know how to gauge the prognosis. In addition, various members of the health care team may offer different opinions. Good communication among providers is essential so that consistent information is provided to patients. This is especially important when the best path forward is uncertain. Advice from experts in palliative and terminal care should be sought whenever appropriate to ensure that clinicians are not providing patients with unrealistic expectations. **For a more complete discussion of end-of-life care, see [Chap. 9](#).**

Maintaining Humanism and Professionalism

Many trends in the delivery of health care tend to make medical care impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but in which the patient may have little choice in selecting a physician; (3) increasing reliance on technological advances and computerization; and (4) the need for numerous physicians and other health professionals to be involved in the care of most patients who are seriously ill.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. While medical schools appropriately place substantial emphasis on professionalism, a physician's personal attributes, including integrity, respect, and compassion, also are extremely important. In the United States, the Gold Humanism Society recognizes individuals who are exemplars of humanistic patient care and serve as role models for medical education and training.

Availability to the patient, expression of sincere concern, willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such situations and should consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times.

Another important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through deliberate, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient.

EXPANDING FRONTIERS IN MEDICAL PRACTICE

The Era of "Omics"

In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a wide variety of microbes. The clinical implications of these insights are illustrated by the complete genome sequencing of H1N1 influenza virus in 2009 and the rapid identification of H1N1 influenza as a potentially fatal pandemic illness, leading to the swift development and dissemination of an effective protective vaccine. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, and genotyping is providing a new means to assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the interplay of genetic and environmental factors.

Rapid progress is also being made in other areas of molecular medicine. *Epigenetics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences that influence gene

expression (**Chap. 471**). Every cell of the body has identical DNA sequences; the diverse phenotypes a person's cells manifest are the result of epigenetic regulation of gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. *Proteomics*, the study of the entire library of proteins made in a cell or organ and the complex relationship of these proteins to disease, is enhancing the repertoire of the 23,000 genes in the human genome through alternate splicing, posttranslational processing, and posttranslational modifications that often have unique functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for diagnostic and disease-screening applications. *Microbiomics* is the study of the resident microbes in humans and other mammals, which together compose the microbiome. The human haploid genome has ~23,000 genes, whereas the microbes residing on and in the human body encompass more than 3–4 million genes; these resident microbes are likely to be of great significance with regard to health status. Ongoing research is demonstrating that the microbes inhabiting human mucosal and skin surfaces play a critical role in maturation of the immune system, in metabolic balance, and in disease susceptibility. A variety of environmental factors, including the use and overuse of antibiotics, have been tied experimentally to substantial increases in disorders such as obesity, metabolic syndrome, atherosclerosis, and immune-mediated diseases in both adults and children. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that may be responsible for the lower incidence of asthma among children raised on farms. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of organ dysfunction and disease. It seems likely that disease-associated patterns will be found in lipids, carbohydrates, membranes, mitochondria, and other vital components of cells and tissues. *Exposomics* is the study of the exposome—i.e., the environmental exposures such as smoking, sunlight, diet, exercise, education, and violence that together have an enormous impact on health. All of this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates challenges in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of *systems biology* and *network medicine* are being applied to the enormous body of information now obtainable for every patient and may eventually provide new approaches to classifying disease. **For a more complete discussion of a complex systems approach to human disease, see Chap. 476.**

The rapidity of these advances may seem overwhelming to practicing physicians. However, physicians have an important role to play in ensuring that these powerful technologies and sources of new information are applied judiciously to patient care. Since “omics” are evolving so rapidly, physicians and other health care professionals must engage in continuous learning so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. **For a more complete discussion of genetic testing, see Chap. 457.**

Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. The outbreak of Zika virus infections in the Americas is a cogent example of this phenomenon. In addition, factors such as wars, the migration of refugees, and climate change are contributing to changing disease profiles worldwide. Patients have broader access to unique expertise or clinical trials at distant medical centers, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation—for example, interpretation of radiologic images and pathologic specimens. **For a complete discussion of global issues, see [Chap. 460](#).**

Medicine on the Internet

On the whole, the Internet has had a positive effect on the practice of medicine; through personal computers, a wide range of information is available to physicians and patients almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician regarding the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. By bringing them into direct and timely contact with the latest developments in medical care, this medium also serves to lessen the information gap that has hampered physicians and health care providers in remote areas.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Patients often arrive at a clinic visit with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an “editor” as patients navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Both physicians and patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients and in this regard represents a remarkable resource that was not available to practitioners a generation ago.

Public Expectations and Accountability

The general public’s level of knowledge and sophistication regarding health issues has grown rapidly over the last few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients’ unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care they provide but also for their patients’ satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient's hospitalization if it falls outside certain "average" standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. A growing "pay-for-performance" movement seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. In many parts of the United States, managed (capitated) care contracts with insurers have replaced traditional fee-for-service care, placing the onus of managing the cost of all care directly on the providers and increasing the emphasis on preventive strategies. In addition, physicians are expected to give evidence of their current competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies

The rapid pace of technological advances has profound implications for medical applications that go far beyond the traditional goals of disease prevention, treatment, and cure. Cloning, genetic engineering, gene therapy, human-computer interfaces, nanotechnology, and use of targeted therapies have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment "normal" human performance, replace failing tissues, and substantially prolong life span. Given their unique training, physicians have a responsibility to help shape the debate on the appropriate uses of and limits placed on these new techniques and to consider carefully the ethical issues associated with the implementation of such interventions. As medicine becomes more complex, shared decision-making is increasingly important, particularly in areas such as genetic counseling and end-of-life care, but also in most instances of considering diagnostic and treatment options.

Learning Medicine

More than a century has passed since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. The tools of medicine also change continuously, necessitating formal training in the use of EMRs, large datasets, ultrasound, robotics, and new imaging techniques. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

At a time when the amount of information that must be mastered to practice medicine continues to expand, increasing pressures both within and outside of medicine have led to the implementation of restrictions on the amount of time a physician-in-training can spend in the hospital and in clinics. Because the benefits associated with continuity of medical care and observation of a patient's progress over time were thought to

be outstripped by the stresses imposed on trainees by long hours and by fatigue-related errors, strict limits were set on the number of patients that trainees could be responsible for at one time, the number of new patients they could evaluate in a day on call, and the number of hours they could spend in the hospital. In 1980, residents in medicine worked in the hospital more than 90 hours per week on average. In 1989, their hours were restricted to no more than 80 per week. Resident physicians' hours further decreased by ~10% between 1996 and 2008, and in 2010 the Accreditation Council for Graduate Medical Education further restricted (i.e., to 16 hours per shift) consecutive in-hospital duty hours for first-year residents. The impact of these changes is still being assessed, but the evidence that medical errors have decreased as a consequence is sparse. An unavoidable by-product of fewer hours at the bedside is an increase in the number of "handoffs" of patient responsibility from one physician to another. These transfers often involve a transition from a physician who knows the patient well, having evaluated that individual on admission, to a physician who knows the patient less well. It is imperative that these transitions of responsibility be handled with care and thoroughness, with all relevant information exchanged and acknowledged.

The Physician as Perpetual Student

From the time physicians graduate from medical school, it becomes all too apparent that this milestone is symbolic and that they must embrace the role of a "perpetual student." This realization is at the same time exhilarating and anxiety-provoking. It is exhilarating because physicians can apply constantly expanding knowledge to the treatment of their patients; it is anxiety-provoking because physicians realize that they will never know as much as they want or need to know. Ideally, physicians will translate the latter feeling into energy through which they can continue to improve and reach their potential. It is the physician's responsibility to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen

Being a physician is a privilege. The capacity to apply one's skills for the benefit of fellow human beings is a noble calling. The physician-patient relationship is inherently unbalanced in the distribution of power. In light of their influence, physicians must always be aware of the potential impact of what they do and say and must always strive to strip away individual biases and preferences to find what is best for their patients. To the extent possible, physicians should also act within their communities to promote health and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in taking action to deliver needed care even when personal financial compensation may not be available.

Research, Teaching, and the Practice of Medicine

The word *doctor* is derived from the Latin *docere*, "to teach." As teachers, physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of their broader societal responsibilities, physicians should encourage patients to participate in ethical and properly approved clinical investigations if

these studies do not impose undue hazard, discomfort, or inconvenience. Physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients. The best interests of the patient must always take priority.

To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

—William Osler, 1849–1919

FURTHER READING

Cheston CC et al: Social media use in medical education: A systematic review. Acad Med 88:893, 2013. [\[PubMed: 23619071\]](#)

Cooke M et al: American medical education 100 years after the Flexner report. N Engl J Med 355:1339, 2006. [\[PubMed: 17005951\]](#)

Institute of Medicine: Dying in America: improving quality and honoring individual preferences near the end of life. Washington, DC, National Academies Press, 2015.

Institute of Medicine: Improving diagnosis in health care. Washington, DC, National Academies of Sciences, Engineering, and Medicine, 2015.

Jameson JL, Longo DL: Precision medicine—personalized, problematic, and promising. N Engl J Med 372:2229, 2015. [\[PubMed: 26014593\]](#)

Reeves S et al: Interprofessional education: Effects on professional practice and healthcare outcomes (update). Cochrane Database Syst Rev 28:CD02213, 2013.

Stern DT, Papadakis M: The developing physician—becoming a professional. N Engl J Med 355:1794, 2006. [\[PubMed: 17065641\]](#)

Vickrey BG et al: How neurologists think: A cognitive psychology perspective on missed diagnoses. Ann Neurol 67:425, 2010. [\[PubMed: 20437577\]](#)

West P et al: Intervention to promote physician well-being, job satisfaction, and professionalism. A randomized clinical trial. JAMA Intern Med 174:527, 2014. [\[PubMed: 24515493\]](#)

3/14/2018

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 56: Cutaneous Drug Reactions

Robert Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai

INTRODUCTION

Cutaneous reactions are among the most frequent adverse reactions to drugs. Most are benign, but a few can be life threatening. Prompt recognition of severe reactions, drug withdrawal, and appropriate therapeutic interventions can minimize toxicity. This chapter focuses on adverse cutaneous reactions to systemic medications; it covers their incidence, patterns, and pathogenesis, and provides some practical guidelines on treatment, assessment of causality, and future use of drugs.

USE OF PRESCRIPTION DRUGS IN THE UNITED STATES

In the United States, more than 3 billion prescriptions for >60,000 drug products, which include >2000 different active agents, are dispensed annually. Hospital inpatients alone annually receive about 120 million courses of drug therapy, and half of adult Americans receive prescription drugs on a regular outpatient basis. Adverse effects of a prescription medication may result in 4.5 million urgent or emergency care visits each year in the United States. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several large cohort studies established that acute cutaneous reactions to drugs affect about 3% of hospitalized patients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

Many drugs of common use are associated with a 1–2% rate of rashes during premarketing clinical trials. The risk is often higher when medications are used in general, unselected populations. The rate may reach 3–7% for amoxicillin, sulfamethoxazole, many anticonvulsants, and anti-HIV agents.

In addition to acute eruptions, a variety of skin diseases can be induced or exacerbated by prolonged use of drugs (e.g., pruritus, pigmentation, nail or hair disorders, psoriasis, bullous pemphigoid, photosensitivity, and even cutaneous neoplasms). These drug reactions are not frequent, but neither their incidence nor their impact on public health has been evaluated.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions. Severe reactions are too rare to be detected in such cohorts. Although rare,

severe cutaneous reactions to drugs have an important impact on health because of significant sequelae, including mortality. Adverse drug rashes are responsible for hospitalization, increase the duration of hospital stay, and can be life threatening. Some populations are at increased risk of drug reactions, including elderly patients, patients with autoimmune disease, hematopoietic stem cell transplant recipients, and those with acute Epstein-Barr virus (EBV) or human immunodeficiency virus (HIV) infection. The pathophysiology underlying this association is unknown but may be related to immunocompromise or immune dysregulation. Individuals with advanced HIV disease (e.g., CD4 T lymphocyte count <200 cells/ μ L) have a 40- to 50-fold increased risk of adverse reactions to sulfamethoxazole ([Chap. 197](#)) and increased risk of severe hypersensitivity reactions.

PATHOGENESIS OF DRUG REACTIONS

Adverse cutaneous responses to drugs can arise as a result of immunologic or nonimmunologic mechanisms.

NONIMMUNOLOGIC DRUG REACTIONS

Examples of nonimmunologic drug reactions are pigmentary changes due to dermal accumulation of medications or their metabolites, alteration of hair follicles by antimetabolites and signaling inhibitors, and lipodystrophy associated with metabolic effects of anti-HIV medications. These side effects are predictable and sometimes can be prevented.

IMMUNOLOGIC DRUG REACTIONS

Evidence suggests an immunologic basis for most acute drug eruptions. Drug reactions may result from immediate release of preformed mediators (e.g., urticaria, anaphylaxis), antibody-mediated reactions, immune complex deposition, and antigen-specific responses. Drug-specific T cell clones can be derived from the blood or from skin lesions of patients with a variety of drug allergies, strongly suggesting that these T cells mediate drug allergy in an antigen-specific manner. Specific clones are generated by medications that are frequently a cause of drug eruptions: penicillin G, amoxicillin, cephalosporins, sulfamethoxazole, phenobarbital, carbamazepine, and [lamotrigine](#). Both CD4 and CD8 clones have been obtained; however, their specific roles in drug allergy have not been elucidated. Drug presentation to T cells is major histocompatibility complex (MHC)-restricted and likely involves drug-peptide complex recognition by specific T cell receptors (TCRs).

Once a drug has induced an immune response, the final phenotype of the reaction is determined by the nature of effectors: cytotoxic (CD8+) T cells in blistering and certain hypersensitivity reactions, chemokines for reactions mediated by neutrophils or eosinophils, and B cell collaboration for production of specific antibodies for urticarial reactions. Immunologic reactions have recently been classified into further subtypes that provide a useful framework for designating adverse drug reactions based on involvement of specific immune pathways ([Table 56-1](#)).

TABLE 56-1

Classification of Adverse Drug Reactions Based on Immune Pathway

Type	Key Pathway	Key Immune Mediators	Adverse Drug Reaction Type
Type I	IgE	IgE	Urticaria, angioedema, anaphylaxis
Type II	IgG-mediated cytotoxicity	IgG	Drug-induced hemolysis, thrombocytopenia (e.g., penicillin)
Type III	Immune complex	IgG + antigen	Vasculitis, serum sickness, drug-induced lupus
Type IVa	T lymphocyte-mediated macrophage inflammation	IFN- γ , TNF- α T _H 1 cells	Tuberculin skin test, contact dermatitis
Type IVb	T lymphocyte-mediated eosinophil inflammation	IL-4, IL-5, IL-13 T _H 2 cells Eosinophils	DIHS Morbilliform eruption
Type IVc	T lymphocyte-mediated cytotoxic T lymphocyte inflammation	Cytotoxic T lymphocytes Granzyme Perforin Granulysin (SJS/TEN] only)	SJS/TEN Morbilliform eruption
Type IVd	T lymphocyte-mediated neutrophil inflammation	CXCL8, IL-17, GM-CSF Neutrophils	AGEP

Abbreviations: AGEP, acute generalized exanthematous pustulosis; DIHS, drug-induced hypersensitivity syndrome; GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis; TNF, tumor necrosis factor.

Immediate Reactions

Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophils. These mediators include histamine, leukotrienes, prostaglandins, bradykinins, platelet-activating factor, enzymes, and proteoglycans. Drugs can trigger mediator release either directly (“anaphylactoid” reaction) or through IgE-specific antibodies. These reactions usually manifest in the skin and gastrointestinal, respiratory, and cardiovascular systems ([Chap. 346](#)). Primary symptoms and signs include pruritus, urticaria, nausea, vomiting, abdominal cramps, bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. Nonsteroidal anti-inflammatory drugs (NSAIDs), including aspirin, and radiocontrast media are frequent causes of direct mast cell degranulation or anaphylactoid reactions, which can occur on first exposure. Penicillins and muscle relaxants used in general anesthesia are the most frequent causes of IgE-dependent reactions to drugs, which require prior sensitization. Release of mediators is triggered when polyvalent drug protein conjugates cross-link IgE molecules fixed to sensitized cells. Certain routes of administration favor different clinical patterns (e.g., gastrointestinal effects from oral route, circulatory effects from intravenous route).

Immune Complex–Dependent Reactions

Serum sickness is produced by tissue deposition of circulating immune complexes with consumption of complement. It is characterized by fever, arthritis, nephritis, neuritis, edema, and an urticarial, papular, or purpuric rash ([Chap. 356](#)). First described following administration of nonhuman sera, it currently occurs in the setting of monoclonal antibodies and similar medications. In classic serum sickness, symptoms develop 6 or more days after drug exposure, the latent period representing the time needed to synthesize antibody. Vasculitis, a relatively rare complication of drugs, may also be a result of immune complex deposition ([Chap. 356](#)). Cephalosporin and other medications, including monoclonal antibodies such as infliximab, rituximab, and [omalizumab](#), may be associated with clinically similar “serum sickness–like” reactions. The mechanism of this reaction is unknown but is unrelated to immune complex formation and complement activation.

Delayed Hypersensitivity

While not completely understood, delayed hypersensitivity directed by drug-specific T cells is an important mechanism underlying the most common drug eruptions, that is, morbilliform eruptions, and also rare and severe forms such as drug-induced hypersensitivity syndrome (DIHS) (also known as drug rash with eosinophilia and systemic symptoms [DRESS]), acute generalized exanthematous pustulosis (AGEP), Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) ([Table 56-1](#)). Drug-specific T cells have been detected in these types of drug eruptions. In TEN, skin lesions contain T lymphocytes reactive to autologous lymphocytes and keratinocytes in a drug-specific, human leukocyte antigen (HLA)-restricted, and perforin/granzyme-mediated pathway.

The mechanism(s) by which medications result in T cell activation is unknown. Two hypotheses prevail: first, that the antigens driving these reactions may be the native drug itself or components of the drug covalently complexed with endogenous proteins, presented in association with HLA molecules to T cells through the classic antigen presentation pathway or, alternatively, through direct interaction of the drug/metabolite with the TCR or peptide-loaded HLA (e.g., the pharmacologic interaction of drugs with immune receptors, or p-i

hypothesis). Recent x-ray crystallography data characterizing binding between specific HLA molecules to particular drugs known to cause hypersensitivity reactions demonstrate unique alterations to the MHC peptide-binding groove, suggesting a molecular basis for T cell activation in the development of hypersensitivity reactions.

GENETIC FACTORS AND CUTANEOUS DRUG REACTIONS



Genetic determinants may predispose individuals to severe drug reactions by affecting either drug metabolism or immune responses to drugs. Polymorphisms in cytochrome P450 enzymes, drug acetylation, methylation (such as thiopurine methyltransferase activity and azathioprine), and other forms of metabolism (such as glucose-6-phosphate dehydrogenase and [dapsons](#)) may increase susceptibility to drug toxicity or underdosing, highlighting a role for differential pharmacokinetic or pharmacodynamic effects. The value of routine screening of P450 enzymes has not been determined, though its cost-effectiveness in certain populations (e.g., patients with seizure disorder) has been suggested.

Associations between drug hypersensitivities and HLA haplotypes suggest a key role for immune mechanisms. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B^{*}57:01 ([Chap. 197](#)). In Taiwan, within a homogeneous Han Chinese population, a 100% association was observed between SJS/TEN (but not DIHS) related to carbamazepine and HLA-B^{*}15:02. In the same population, another 100% association was found between HLA-B^{*}58:01 and SJS, TEN, or DIHS related to [allopurinol](#). These associations are drug and phenotype specific; that is, HLA-specific T cell stimulation by medications leads to distinct reactions. However, the strong associations found in Taiwan have not been observed in other countries with more heterogeneous populations.

GLOBAL CONSIDERATIONS



Recognition of HLA associations with drug hypersensitivity has resulted in recommendations to screen high-risk populations. Genetic screening for HLA-B^{*}57:01 to prevent abacavir hypersensitivity, which carries a 100% negative predictive value when patch test confirmed and 55% positive predictive value generalizable across races, is becoming the clinical standard of care worldwide (number needed to treat = 13). The U.S. Food and Drug Administration has recommended HLA-B^{*}15:02 screening of Asian individuals prior to a new prescription of carbamazepine. The American College of Rheumatology has recommended HLA-B^{*}58:01 screening of Han Chinese patients prescribed [allopurinol](#). To date, screening for a single HLA (but not multiple HLA haplotypes) in specific populations has been determined to be cost-effective.

Several investigators have proposed that specific HLA haplotypes associated with drug hypersensitivity indeed play a pathogenic role; stimulation of carbamazepine-specific cytotoxic T lymphocytes (CTLs) in the context of HLA-B^{*}15:02 results in production of a putative mediator of keratinocyte necrosis in TEN. Other studies have identified CTLs reactive to carbamazepine that use highly restricted V-alpha and V-beta TCR

repertoires in patients with carbamazepine hypersensitivity that are not found in carbamazepine-tolerant individuals. Genetic testing for specific HLA haplotypes and functional screening for TCR repertoire to identify patients at risk is becoming more widely available and heralds the era of personalized medicine and pharmacogenomics.

CLINICAL PRESENTATION OF CUTANEOUS DRUG REACTIONS

NONIMMUNE CUTANEOUS REACTIONS

Exacerbation or Induction of Dermatologic Diseases

A variety of drugs can exacerbate preexisting diseases or induce—or unmask—a disease that may or may not disappear after withdrawal of the inducing medication. For example, NSAIDs, lithium, beta blockers, tumor necrosis factor (TNF) antagonists, interferon (IFN) α , and angiotensin-converting enzyme (ACE) inhibitors can exacerbate plaque psoriasis, whereas antimalarials and withdrawal of systemic glucocorticoids can worsen pustular psoriasis. The situation of TNF- α inhibitors is unusual, as this class of medications is used to treat psoriasis; however, they may induce psoriasis (especially palmoplantar) in patients being treated for other conditions. Acne may be induced by glucocorticoids, androgens, lithium, and antidepressants. Follicular papular or pustular eruptions of the face and trunk resembling acne frequently occur with epidermal growth factor receptor (EGFR) antagonists. The severity of the eruption correlates with a better anticancer effect. This rash is typically responsive to and prevented by tetracycline antibiotics.

Several medications induce or exacerbate autoimmune disease. Interleukin (IL) 2, IFN- α , and anti-TNF- α are associated with new-onset systemic lupus erythematosus (SLE). Drug-induced lupus is classically marked by antinuclear and antihistone antibodies and, in some cases, anti-double-stranded DNA (D-penicillamine, anti-TNF- α) or perinuclear anti-neutrophil cytoplasmic antibody (p-ANCA) (minocycline) antibodies. Subacute lupus erythematosus (SCLE) can be induced by a growing list of drugs, including thiazide diuretics, TNF-inhibitors, [terbinafine](#), and minocycline. IFN and TNF-inhibitors can induce granulomatous disease and sarcoidosis. Autoimmune blistering diseases may be drug induced as well: pemphigus by D-penicillamine and ACE inhibitors, bullous pemphigoid by [furosemide](#) and PD-1 inhibitors, and linear IgA bullous dermatosis by vancomycin. Other medications may cause highly specific cutaneous reactions. Gadolinium contrast has been associated with nephrogenic systemic fibrosis, a condition of sclerosing skin with rare internal organ involvement; advanced renal compromise may be an important risk factor. Granulocyte colony-stimulating factor, azacitidine, all-trans retinoic acid, and the *FLT3*-inhibitor class of drugs may induce neutrophilic dermatoses. In this setting, the hypothesis that a drug may be responsible should always be considered, even after the treatment is complete. In addition, reactions may develop in cases of long-term medication therapy due to small changes in dosing or host metabolism. Resolution of the cutaneous reaction may be delayed upon discontinuation of the medication.

Photosensitivity Eruptions

Photosensitivity eruptions are usually most marked in sun-exposed areas, but they may extend to sun-protected areas. The mechanism is almost always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria, most commonly with NSAIDs. The severity of the reaction depends on the tissue level of the drug, its efficiency as a photosensitizer, and the extent of exposure to the activating wavelengths of ultraviolet (UV) light (**Chap. 57**).

Common orally administered photosensitizing drugs include fluoroquinolones, tetracycline antibiotics, and trimethoprim/sulfamethoxazole. Other drugs less frequently implicated are chlorpromazine, thiazides, NSAIDs, and BRAF inhibitors. Voriconazole may result in severe photosensitivity, accelerated photoaging, and cutaneous carcinogenesis.

Because UV-A and visible light, which trigger these reactions, are not easily absorbed by nonopaque sunscreens and are transmitted through window glass, photosensitivity reactions may be difficult to block. Photosensitivity reactions abate with removal of either the drug or UV radiation, use of sunscreens that block UV-A light, and treatment of the reaction as one would a sunburn. Rarely, individuals develop persistent reactivity to light, necessitating long-term avoidance of sun exposure. Some chemotherapeutic agents, such as methotrexate, can induce a UV-recall reaction characterized by an erythematous, slightly scaly eruption at sites of prior severe sun exposure.

Pigmentation Changes

Drugs, either systemic or topical, may cause a variety of pigmentary changes in the skin by triggering melanocyte production of melanin (as in the case of oral contraceptives causing melasma) or due to deposition of drug or drug metabolites. Long-term minocycline and amiodarone may cause blue-gray pigmentation. Phenothiazine, gold, and bismuth result in gray-brown pigmentation of sun-exposed areas. Numerous cancer chemotherapeutic agents may be associated with characteristic patterns of pigmentation (e.g., bleomycin, busulfan, daunorubicin, cyclophosphamide, hydroxyurea, **fluorouracil**, and methotrexate). **Clofazimine** causes a drug-induced lipofuscinosis with characteristic red-brown coloration.

Hyperpigmentation of the face, mucous membranes, and pretibial and subungual areas occurs with antimalarials. Quinacrine causes generalized yellow discoloration. Pigmentation changes may also occur in mucous membranes (busulfan, bismuth), conjunctiva (chlorpromazine, thioridazine, imipramine, clomipramine), nails (zidovudine, doxorubicin, cyclophosphamide, bleomycin, **fluorouracil**, hydroxyurea), hair, and teeth (tetracyclines).

Warfarin Necrosis of Skin

This rare reaction (0.01–0.1%) usually occurs between the third and tenth days of therapy with warfarin, usually in women. Common sites are breasts, thighs, and buttocks (**Fig. 56-1**). Lesions are sharply demarcated, erythematous, or purpuric, and may progress to form large, hemorrhagic bullae with necrosis and eschar formation.

FIGURE 56-1

Warfarin necrosis involving the breasts.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Warfarin anticoagulation in protein C or S deficiency causes an additional reduction in already low circulating levels of endogenous anticoagulants, permitting hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect areas adjacent to the injection site or more distant sites if infused.

Warfarin-induced cutaneous necrosis is treated with vitamin K, heparin, surgical debridement, and intensive wound care. Treatment with protein C concentrates may also be helpful. Newer anticoagulants such as dabigatran etexilate may avoid warfarin necrosis in high-risk patients.

Drug-Induced Hair Disorders • DRUG-INDUCED HAIR LOSS

Medications may affect hair follicles at two different phases of their growth cycle: anagen (growth) or telogen (resting). *Anagen effluvium* occurs within days of drug administration, especially with antimetabolite or other chemotherapeutic drugs. In contrast, in *telogen effluvium*, the delay is 2–4 months following initiation of a new medication. Both present as diffuse, nonscarring alopecia most often reversible after discontinuation of the responsible agent.

A considerable number of drugs have been associated with hair loss. These include antineoplastic agents (alkylating agents, bleomycin, vinca alkaloids, platinum compounds), anticonvulsants (carbamazepine, valproate), beta blockers, antidepressants, antithyroid drugs, IFNs, oral contraceptives, and cholesterol-lowering agents.

DRUG-INDUCED HAIR GROWTH

Medications may also cause hair growth. Hirsutism is an excessive growth of terminal hair with masculine hair growth pattern in a female, most often on the face and trunk, due to androgenic stimulation of hormone-sensitive hair follicles (anabolic steroids, oral contraceptives, testosterone, corticotropin). Hypertrichosis is a

distinct pattern of hair growth, not in a masculine pattern, typically located on the forehead and temporal regions of the face. Drugs responsible for hypertrichosis include anti-inflammatory drugs, glucocorticoids, vasodilators (diazoxide, [minoxidil](#)), diuretics (acetazolamide), anticonvulsants (phenytoin), immunosuppressive agents ([cyclosporine A](#)), psoralens, and zidovudine.

Changes in hair color or structure are uncommon adverse effects from medications. Hair discoloration may occur with chloroquine, IFN- α , chemotherapeutic agents, and tyrosine kinase inhibitors. Changes in hair structure have been observed in patients given EGFR inhibitors, BRAF inhibitors, tyrosine kinase inhibitors, and acitretin.

Drug-Induced Nail Disorders

Drug-related nail disorders usually involve all 20 nails and need months to resolve after withdrawal of the medication. The pathogenesis is most often toxic. Drug-induced nail changes include Beau's line (transverse depression of the nail plate), onycholysis (detachment of the distal part of the nail plate), onychomadesis (detachment of the proximal part of the nail plate), pigmentation, and paronychia (inflammation of periungual skin).

ONYCHOLYSIS

Onycholysis occurs with tetracyclines, fluoroquinolones, retinoids, NSAIDs, and others, including many chemotherapeutic agents, and may be triggered by exposure to sunlight.

ONYCHOMADESIS

Onychomadesis is caused by temporary arrest of nail matrix mitotic activity. Common drugs reported to induce onychomadesis include carbamazepine, lithium, retinoids, and chemotherapeutic agents.

PARONYCHIA

Paronychia and multiple pyogenic granuloma with progressive and painful periungual abscess of fingers and toes are side effects of systemic retinoids, lamivudine, indinavir, and anti-EGFR monoclonal antibodies.

NAIL DISCOLORATION

Some drugs—including anthracyclines, taxanes, [fluorouracil](#), psoralens, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose dependent.

Toxic Erythema of Chemotherapy and Other Chemotherapy Reactions

Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. A broad spectrum of chemotherapy-related skin toxicities have been reported, including neutrophilic eccrine hidradenitis, sterile cellulitis, exfoliative dermatitis, and flexural erythema; recent nomenclature classifies these under the unifying diagnosis of toxic erythema of chemotherapy (TEC) ([Fig. 56-2](#)). Acral erythema is marked by dysesthesia and an erythematous, edematous eruption of the palms and soles. Common causes include cytarabine, doxorubicin, methotrexate, hydroxyurea, [fluorouracil](#), and capecitabine.

FIGURE 56-2

Toxic erythema of chemotherapy.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

The recent introduction of many new monoclonal antibody and small molecular signaling inhibitors for the treatment of cancer has been accompanied by numerous reports of skin and hair toxicity; only the most common of these are mentioned here. EGFR antagonists induce follicular eruptions and nail toxicity after a mean interval of 10 days in a majority of patients. Xerosis, eczematous eruptions, acneiform eruptions, and pruritus are common. Erlotinib is associated with marked hair textural changes. Sorafenib, a tyrosine kinase inhibitor, may result in follicular eruptions and focal bullous eruptions at palmoplantar, flexural sites or areas of frictional pressure. BRAF inhibitors are associated with photosensitivity, palmoplantar hyperkeratosis, hair curling, dyskeratotic (Grover's-like) rash, hyperkeratotic benign cutaneous neoplasms, and keratoacanthoma-like squamous cell carcinomas. Rash, pruritus, and vitiliginous depigmentation have been reported in association with [ipilimumab](#) (anti-CTLA4) treatment. Up to 50% of patients experience immune-mediated skin eruptions, including granulomatous reactions, dermatomyositis, panniculitis, and vasculitis.

IMMUNE CUTANEOUS REACTIONS: COMMON

Maculopapular Eruptions

Morbilliform or maculopapular eruptions (**Fig. 56-3**) are the most common of all drug-induced reactions, often start on the trunk or intertriginous areas, and consist of blanching erythematous macules and papules that are symmetric and confluent. Nonblanching, dusky, or bright-red macules should raise concern for a more severe reaction. Involvement of mucous membranes is rare and should prompt consideration of SJS. Facial involvement in morbilliform eruptions is also uncommon, and the presence of extensive facial lesions with facial edema suggests DIHS. Diagnosis of morbilliform eruptions is rarely assisted by laboratory testing. Skin biopsy often shows nonspecific inflammatory changes.

FIGURE 56-3

Morbilliform drug eruption.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Morbilliform eruptions may be associated with moderate to severe pruritus and fever. A viral exanthem is another differential diagnostic consideration, especially in children, and graft-versus-host disease is also a consideration in the proper clinical setting. Absence of enanthems; absence of ear, nose, throat, and upper respiratory tract symptoms; and polymorphism of the skin lesions support a drug rather than a viral eruption. Common offenders include aminopenicillins, cephalosporins, antibacterial sulfonamides, [allopurinol](#), and antiepileptic drugs. Beta blockers, calcium channel blockers, and ACE inhibitors are rarely the culprit; however, any drug can cause a morbilliform exanthem. Certain medications carry very high rates of morbilliform eruption, including nevirapine and [lamotrigine](#), even in the absence of DIHS reactions. [Lamotrigine](#) morbilliform rash is associated with higher starting doses, rapid dose escalation, concomitant use of valproate (which increases [lamotrigine](#) levels and half-life), and use in children.

Maculopapular reactions usually develop within 1 week of initiation of therapy and last less than 2 weeks. Occasionally, these eruptions resolve despite continued use of the responsible drug. Because the eruption may also worsen, the suspect drug should be discontinued unless it is essential. It is important to note that

the rash may continue to progress for a few days up to 1 week following medication discontinuation. Oral antihistamines and emollients may help relieve pruritus. Short courses of potent topical glucocorticoids can reduce inflammation and symptoms. Systemic glucocorticoid treatment is rarely indicated.

Pruritus

Pruritus is associated with almost all drug eruptions and, in some cases, may represent the only symptom of the adverse cutaneous reaction. It may be alleviated by antihistamines such as hydroxyzine or [diphenhydramine](#). Pruritus stemming from specific medications may require distinct treatment, such as selective opiate antagonists for opiate-related pruritus.

Urticaria/Angioedema/Anaphylaxis

Urticaria, the second most frequent type of cutaneous reaction to drugs, is characterized by pruritic, red wheals of varying size rarely lasting more than 24 hours. It has been observed in association with nearly all drugs, most frequently ACE inhibitors, aspirin, NSAIDs, penicillin, and blood products. However, medications account for no more than 10–20% of acute urticaria cases. Deep edema within dermal and subcutaneous tissues is known as angioedema and may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Drug-induced urticaria may be caused by three mechanisms: an IgE-dependent mechanism, circulating immune complexes (serum sickness), and nonimmunologic activation of effector pathways. IgE-dependent urticarial reactions usually occur within 36 hours of drug exposure but can occur within minutes. Immune complex–induced urticaria associated with serum sickness–like reactions usually occur 6–12 days after first exposure. In this syndrome, the urticarial eruption (typically polycyclic plaques over distal joints) may be accompanied by fever, hematuria, arthralgias, hepatic dysfunction, and neurologic symptoms. Certain drugs, such as NSAIDs, ACE inhibitors, angiotensin II antagonists, radiographic dye, and opiates, may induce urticarial reactions, angioedema, and anaphylaxis in the absence of drug-specific antibodies through direct mast-cell degranulation.

Radiocontrast agents are a common cause of urticaria and, in rare cases, can cause anaphylaxis. High-osmolality radiocontrast media are about five times more likely to induce urticaria (1%) or anaphylaxis than are newer low-osmolality media. About one-third of those with mild reactions to previous exposure react on reexposure. Pretreatment with [prednisone](#) and [diphenhydramine](#) reduces reaction rates.

The treatment of urticaria or angioedema depends on the severity of the reaction. In severe cases with respiratory or cardiovascular compromise, [epinephrine](#) and intravenous glucocorticoids are the mainstay of therapy. For patients with urticaria without symptoms of angioedema or anaphylaxis, drug withdrawal and oral antihistamines are usually sufficient. Future drug avoidance is recommended; rechallenge, especially in individuals with severe reactions, should only occur in an intensive care setting.

Anaphylactoid Reactions

Vancomycin is associated with red man syndrome, a histamine-related anaphylactoid reaction characterized by flushing, diffuse maculopapular eruption, and hypotension. In rare cases, cardiac arrest may be associated with rapid IV infusion of the medication.

Irritant/Allergic Contact Dermatitis

Patients using topical medications may develop an irritant or allergic contact dermatitis to the medication itself or to a preservative or other component of the formulation. Reactions to neomycin sulfate, **bacitracin**, and **polymyxin B** are common. Contact dermatitis may be seen to adhesive tapes, leading to irritation or blisters around ports and IV sites (**Fig. 56-4**). Harsh disinfectant skin cleansers may lead to localized irritant dermatitis.

FIGURE 56-4

Allergic contact dermatitis (bullous) due to adhesive tape.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Fixed Drug Eruptions

These less common reactions are characterized by one or more sharply demarcated, dull red to brown lesions, sometimes with central dusky violaceous erythema and central bulla (**Fig. 56-5**). Hyperpigmentation often results after resolution of the acute inflammation. With rechallenge, the process recurs in the same (fixed) location but may spread to new areas as well. Lesions often involve the lips, hands, legs, face, genitalia, and oral mucosa, and cause a burning sensation. Most patients have multiple lesions. Fixed drug

eruptions have been associated with pseudoephedrine (frequently a nonpigmenting reaction), phenolphthalein (in laxatives), sulfonamides, tetracyclines, NSAIDs, barbiturates, and others.

FIGURE 56-5

Fixed drug eruption.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

IMMUNE CUTANEOUS REACTIONS: RARE AND SEVERE

Drug-Induced Hypersensitivity Syndrome

DIHS is a systemic drug reaction also known as DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome; since eosinophilia is not always present, the term *DIHS* is now preferred. Clinically, DIHS presents with a prodrome of fever and flu-like symptoms for several days, followed by the appearance of a diffuse morbilliform eruption usually involving the face (**Fig. 56-6**). Facial swelling and hand/foot swelling are often present. Systemic manifestations include lymphadenopathy, fever, and leukocytosis (often with eosinophilia or atypical lymphocytosis), as well as hepatitis, nephritis, pneumonitis, myositis, and gastroenteritis, in descending order. Distinct patterns of timing of onset and organ involvement may exist; for example [allopurinol](#) classically induces DIHS with renal involvement, cardiac and lung involvements are more common with minocycline, gastrointestinal involvement is almost exclusively seen with abacavir, and some medications typically lack eosinophilia (abacavir, [dapson](#), [lamotrigine](#)). The cutaneous reaction usually begins 2–8 weeks after the drug is started and persists after drug cessation. Signs and symptoms may continue for several weeks, especially those associated with hepatitis. The eruption recurs with rechallenge, and cross-reactions among aromatic anticonvulsants, including phenytoin, carbamazepine, and

phenobarbital, are common. Other drugs causing DIHS include antibacterial sulfonamides and other antibiotics. Hypersensitivity to reactive drug metabolites, hydroxylamine for sulfamethoxazole and arene oxide for aromatic anticonvulsants, may be involved in the pathogenesis of DIHS. Reactivation of herpes viruses, in particular human herpesviruses 6 and 7, EBV, and cytomegalovirus (CMV), has been frequently reported in this syndrome, although the causal role of viral infection has been debated. Recent research suggests that inciting drugs may reactivate quiescent herpes viruses, resulting in expansion of viral-specific CD8+ T lymphocytes and subsequent end-organ damage. Viral reactivation may be associated with a worse clinical prognosis. Mortality rates as high as 10% have been reported, with most fatalities resulting from liver failure. Systemic glucocorticoids (1.5–2 mg/kg/d [prednisone](#) equivalent) should be started and tapered slowly over 8–12 weeks, during which time clinical symptoms and labs (including complete blood count with differential, basic metabolic panel, and liver function tests) should be followed carefully. A steroid-sparing agent such as mycophenolate mofetil may be indicated in cases of rapid recurrence upon steroid taper. In all cases, immediate withdrawal of the suspected culprit drug is required. Given the severe long-term complications of myocarditis, patients should undergo cardiac evaluation in cases of severe DIHS or if heart involvement is suspected due to hypotension or arrhythmia. Patients should be closely monitored for resolution of organ dysfunction and for development of late-onset autoimmune thyroiditis and diabetes (up to 6 months).

FIGURE 56-6

Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis

SJS and *TEN* are characterized by blisters and mucosal/epidermal detachment resulting from full-thickness epidermal necrosis in the absence of substantial dermal inflammation. The term *Stevens-Johnson syndrome* (*SJS*) describes cases in which the total body surface area of blistering and eventual detachment is <10% (**Fig. 56-7**). The term *Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) overlap* is used to describe cases with 10–30% epidermal detachment (**Fig. 56-8**), and *TEN* is used to describe cases with >30% detachment (**Figs. 56-9 and 56-10**).

FIGURE 56-7

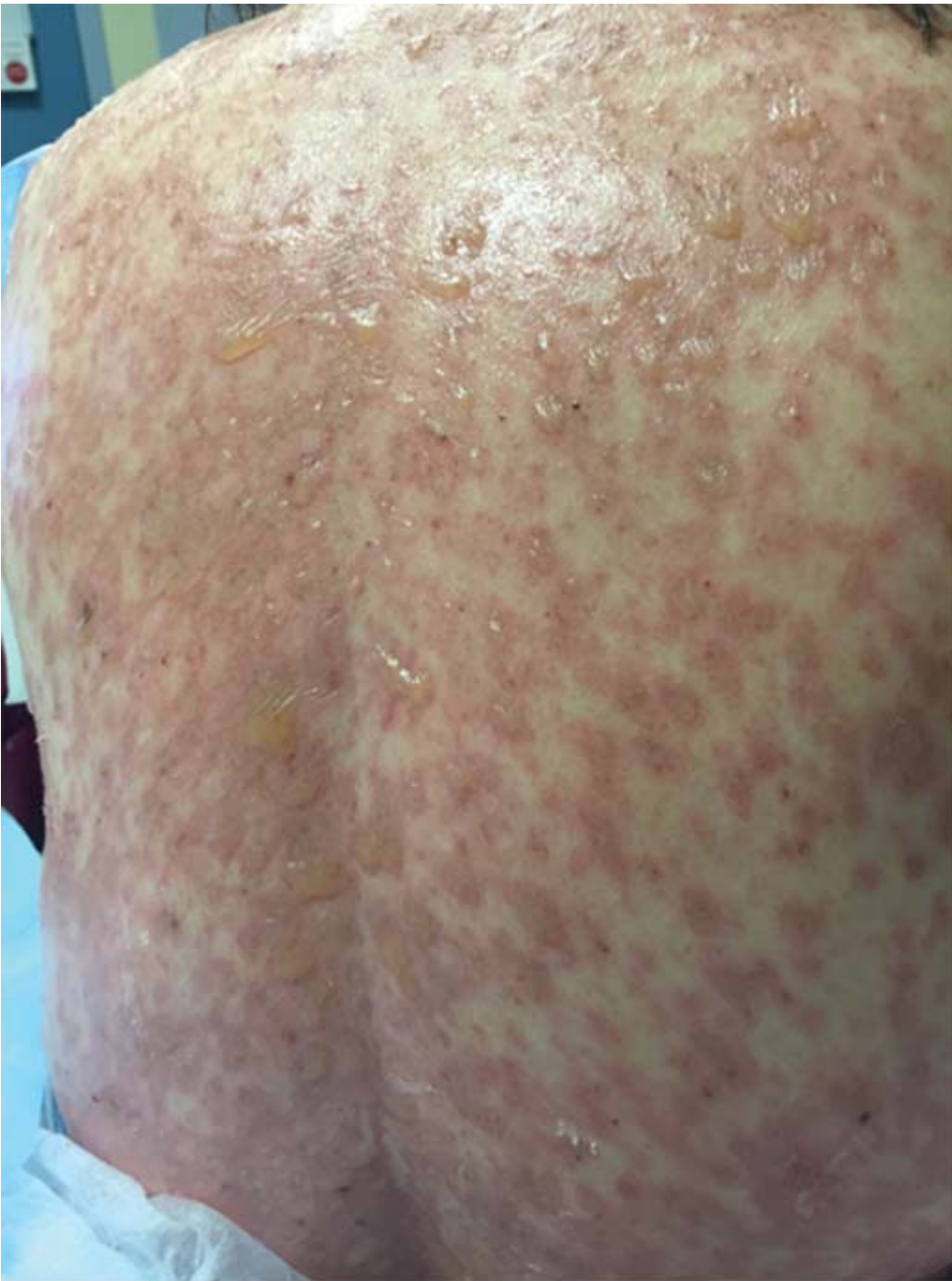
Stevens-Johnson syndrome (SJS).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 56-8

SJS-TEN overlap.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 56-9

Toxic epidermal necrolysis, hand.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 56-10

Toxic epidermal necrolysis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Other blistering eruptions with concomitant mucositis may be confused with SJS/TEN. Erythema multiforme (EM) associated with herpes simplex virus is characterized by painful mucosal erosions and target lesions, typically with an acral distribution and limited skin detachment. *Mycoplasma* infection in children causes a clinically distinct presentation with prominent mucositis and limited cutaneous involvement. The name *Mycoplasma*-induced rash and mucositis has been proposed to help differentiate this clinical entity, which some believe may be the syndrome originally described by Stevens and Johnson.

Patients with SJS/TEN initially present with fever $>39^{\circ}\text{C}$ (102.2°F); sore throat; conjunctivitis; and acute onset of painful dusky, atypical, target-like lesions (**Fig. 56-11**). Intestinal and upper respiratory tract involvement are associated with a poor prognosis, as are older age and greater extent of epidermal detachment. At least 10% of those with SJS and 30% of those with TEN die from the disease. Drugs that most commonly cause SJS/TEN are sulfonamides, [allopurinol](#), antiepileptics (e.g., [lamotrigine](#), phenytoin, carbamazepine), oxicam NSAIDs, β -lactam and other antibiotics, and nevirapine. Frozen-section skin biopsy may aid in rapid diagnosis. At this time, there is no consensus on the most effective treatment for SJS/TEN. The best outcomes stem from early diagnosis, immediate discontinuation of the suspected drug, and meticulous

supportive therapy in an intensive care or burn unit. Issues such as fluid management, atraumatic wound care, infection prevention and treatment, and ophthalmologic and respiratory support are critical. Systemic glucocorticoid therapy ([prednisone](#) 1–2 mg/kg) may be useful early in disease evolution; however, long-term or late systemic glucocorticoid use has been associated with increased mortality. After initial enthusiasm for the use of intravenous immunoglobulin (IVIG) in the treatment of SJS/TEN, more recent data question whether it is beneficial. There are emerging data to support treatment with [cyclosporine](#) and etanercept. Randomized studies to evaluate potential therapies are lacking and difficult to perform.

FIGURE 56-11

Target-like lesion in SJS.

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Pustular Eruptions

AGEP is a rare reaction pattern affecting 3–5 people per million per year. It is thought to be secondary to medication exposure in >90% of cases ([Fig. 56-12](#)). Patients typically present with diffuse erythema or erythroderma, as well as high spiking fevers, and leukocytosis. One to two days later, innumerable pinpoint pustules develop overlying the erythema. The pustules are most pronounced in body fold areas; however, they may become generalized and, when coalescent, can lead to superficial erosion. In such cases, differentiating the eruption from SJS in its initial stages may be difficult; in AGEP, any erosions tend to be more superficial, and prominent mucosal involvement is lacking. Skin biopsy shows collections of neutrophils and sparse necrotic keratinocytes in the upper part of the epidermis, unlike the full-thickness

epidermal necrosis that characterizes SJS. Before the pustules appear, AGEP may also mimic DIHS due to the prominent fever and erythroderma.

FIGURE 56-12

Acute generalized exanthematous pustulosis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

The principal differential diagnosis for AGEP is acute pustular psoriasis, which has an identical clinical and histologic appearance. Many patients with AGEP have a personal or family history of psoriasis. AGEP classically begins within 24–48 hours of drug exposure, though it may occur as much as 1–2 weeks later. β -Lactam antibiotics, calcium channel blockers, macrolide antibiotics, and other inciting agents (including radiocontrast and dialysates) have been reported. Patch testing with the responsible drug often results in a localized pustular eruption.

Overlap Hypersensitivity Syndromes

An important concept in the clinical approach to severe drug eruptions is the presence of overlap syndromes, most notably DIHS with TEN-like features, DIHS with pustular eruption (AGEP-like), and AGEP with TEN-like features. In several case series of AGEP, 50% of cases had TEN-like or DRESS-like features, and 20% of cases had mucosal involvement resembling SJS/TEN. In one study, up to 20% of all severe drug eruptions had overlap features, suggesting that AGEP, DIHS, and SJS/TEN represent a clinical spectrum with some common pathophysiologic mechanisms. Designation of a single diagnosis based on cutaneous and extracutaneous involvement may not always be possible in cases of hypersensitivity; in such instances, treatment should be geared toward addressing the dominant clinical features. The timing of rash onset with respect to drug

administration, which is usually much more delayed in DIHS, and the presence of systemic manifestations such as hepatitis are helpful clues to that diagnosis.

Vasculitis

Cutaneous small-vessel vasculitis (CSVV) typically presents with purpuric papules and macules involving the lower extremities and other dependent areas (**Fig. 56-13**) (**Chap. 356**). Pustular and hemorrhagic vesicles as well as rounded ulcers also occur. Importantly, vasculitis may involve other organs, including the kidneys, joints, gastrointestinal tract, and lungs, necessitating a thorough clinical evaluation for systemic involvement. Drugs are implicated as a cause of roughly 15% of all cases of small vessel vasculitis. Antibiotics, particularly β -lactams, are commonly implicated; however, almost any drug can cause vasculitis. Vasculitis may also be idiopathic or due to underlying infection, connective tissue disease, or (rarely) malignancy.

FIGURE 56-13

Cutaneous small-vessel vasculitis (CSVV, leukocytoclastic vasculitis).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Rare but important types of drug-induced vasculitis include drug-induced ANCA vasculitis. Such patients commonly present with cutaneous manifestations but can develop the full range of symptoms associated with ANCA vasculitis, including crescentic glomerulonephritis and alveolar hemorrhage. Propylthiouracil, methimazole, and hydralazine are common culprits. Drug-induced polyarteritis nodosa has been associated with long-term exposure to minocycline. The presence of perivascular eosinophils on skin biopsy can be a clue to possible drug etiology.

MANAGEMENT OF THE PATIENT WITH SUSPECTED DRUG ERUPTION

There are four main questions to answer regarding a suspected drug eruption:

1. Is the observed rash caused by a medication?
2. Is the reaction severe or evolving?
3. Which drug or drugs are suspected, and should they be withdrawn?
4. What recommendation can be made for future medication use?

EARLY DIAGNOSIS OF SEVERE ERUPTIONS

Rapid recognition of potentially serious or life-threatening reactions is paramount. In this regard, a suspected drug eruption is best defined initially by what it is not (e.g., SJS/TEN, DIHS). [Table 56-2](#) lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. [Table 56-3](#) lists the most important of these reactions, along with their key features and commonly associated medications. Any concern for a serious reaction should prompt immediate consultation with a dermatologist and/or referral of the patient to a specialized center.

TABLE 56-2

Clinical and Laboratory Findings Suggestive of Severe Cutaneous Adverse Drug Reaction

Cutaneous
Generalized erythema Facial edema Skin pain Palpable purpura Dusky or target-like lesions Skin necrosis Blisters or epidermal detachment Positive Nikolsky sign Mucous membrane erosions Swelling of lips or tongue
General
High fever Enlarged lymph nodes Arthralgias or arthritis Shortness of breath, hoarseness, wheezing, hypotension
Laboratory Results
Eosinophil count >1000/ μ L Lymphocytosis with atypical lymphocytes Abnormal liver or kidney function tests

Source: Adapted from Roujeau JC, Stern RS: Severe adverse cutaneous reactions to drugs. N Engl J Med 331:1272, 1994.

TABLE 56-3

Clinical Features of Severe Cutaneous Drug Reactions

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Stevens-Johnson syndrome (SJS)	Erosions usually at two or more sites	Small blisters form from dusky macules or atypical targets; rare areas of confluence; detachment $\leq 10\%$ body surface area	Most cases involve fever	Sulfonamides, anticonvulsants, allopurinol , nonsteroidal anti-inflammatory drugs (NSAIDs)
Toxic epidermal necrolysis (TEN) ^a	Erosions usually at two or more sites	Individual lesions like those seen in SJS; confluent dusky erythema; large sheets of necrotic epidermis; total detachment of $>30\%$ body surface area	Nearly all cases involve fever, “acute skin failure,” leukopenia	Same as for SJS
Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS)	Mucositis reported in as many as 30%	Diffuse, deep red morbilliform eruption with facial involvement; facial and acral swelling	Fever, lymphadenopathy, hepatitis, nephritis, myocarditis, eosinophilia, atypical lymphocytosis	Anticonvulsants, sulfonamides, allopurinol , minocycline
Acute generalized exanthematous pustulosis (AGEP)	Oral erosions in perhaps 20%	Innumerable pinpoint pustules overlying a diffuse erythematous eruption; may develop superficial erosions	High fever, leukocytosis (neutrophilia), hypocalcemia	β -Lactam antibiotics, calcium channel blockers, macrolide antibiotics

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Serum sickness or serum sickness-like reaction	Absent	Urticarial serpiginous or polycyclic rash; purpuric eruption along the sides of the feet and hands is characteristic	Fever, arthralgias	Antithymocyte globulin, cephalosporins, monoclonal antibodies
Anticoagulant-induced necrosis	Infrequent	Purpura and necrosis, especially of central, fatty areas	Pain in affected areas	Warfarin, heparin
Angioedema	Often involved	Urticaria or swelling of the central face, other areas	Respiratory distress, cardiovascular collapse	Angiotensin-converting enzyme (ACE) inhibitors, NSAIDs, contrast dye

^aOverlap of SJS and TEN have features of both, and attachment of 10–30% of body surface area may occur.

Source: Adapted from Roujeau JC, Stern RS: Severe adverse cutaneous reactions to drugs. N Engl J Med 331:1272, 1994.

CONFIRMATION OF DRUG REACTION

The probability of drug etiology varies with the pattern of the reaction. Only fixed drug eruptions are always drug-induced. Morbilliform eruptions are usually viral in children and drug-induced in adults. Among severe reactions, drugs account for 10–20% of anaphylaxis and vasculitis and between 70% and 90% of AGEP, DIHS, SJS, and TEN. Skin biopsy helps characterize the reaction but does not indicate drug causality. Blood counts and liver and renal function tests are important for evaluating organ involvement. The association of mild elevation of liver enzymes and high eosinophil count is frequent but not specific for a drug reaction. Blood tests that could identify an alternative cause, serologic tests (to rule out drug-induced lupus), and serology or polymerase chain reaction for infections may be of great importance to determine a cause.

WHAT DRUG(S) TO SUSPECT AND WITHDRAW

Most cases of drug eruptions occur during the first course of treatment with a new medication. A notable exception is IgE-mediated urticaria and anaphylaxis that need presensitization and develop a few minutes to a few hours after rechallenge. Characteristic timing of onset following drug administration is as follows: 4–14

days for morbilliform eruption, 2–4 days for AGEP, 5–28 days for SJS/TEN, and 14–48 days for DIHS. A drug chart, compiling information of all current and past medications/supplements and the timing of administration relative to the rash, is a key diagnostic tool for identifying the inciting drug. Medications introduced for the first time in the relevant time frame are prime suspects. Two other important elements to suspect causality at this stage are (1) previous experience with the drug in the population and (2) alternative etiologic candidates.

The decision to continue or discontinue any medication depends on the severity of the reaction, the severity of the primary disease undergoing treatment, the degree of suspicion of causality, and the feasibility of finding an alternative safer treatment. In any potentially fatal drug reaction, elimination of all possible suspect drugs or unnecessary medications should be immediately attempted. Some rashes may resolve when “treating through” a benign drug-related eruption. The decision to treat through an eruption should, however, remain the exception and withdrawal of every suspect drug the general rule. On the other hand, drugs that are not suspected and are important for the patient (e.g., antihypertensive agents) generally should not be quickly withdrawn. This approach may permit judicious use of these agents in the future.

RECOMMENDATION FOR FUTURE USE OF DRUGS

The aims are to (1) prevent the recurrence of the drug eruption and (2) avoid compromising future treatment by inaccurately excluding otherwise useful medications.

A thorough assessment of drug causality is based on timing of the reaction, evaluation of other possible causes, and effect of drug withdrawal or continuation. The RegiSCAR group has proposed the Algorithm of Drug Causality for Epidermal Necrolysis (ALDEN) to rank likelihood of drug causality in SJS/TEN; validation of this and other instruments, such as the Naranjo adverse drug reaction probability scale, is limited.

Medication(s) with a “definite” or “probable” causality should be contraindicated, a warning card or medical alert tag (e.g., wristband) should be given to the patient, and the drugs should be listed in the patient’s medical chart as allergies.

CROSS-SENSITIVITY

Because of possible cross-sensitivity among chemically related drugs, many physicians recommend avoidance of not only the medication that induced the reaction but also all drugs of the same pharmacologic class.

There are two types of cross-sensitivity. Reactions that depend on a pharmacologic interaction may occur with all drugs that target the same pathway, whether the drugs are structurally similar or not. This is the case with angioedema caused by NSAIDs and ACE inhibitors. In this situation, the risk of recurrence varies from drug to drug in a particular class; however, avoidance of all drugs in the class is usually recommended.

Immune recognition of structurally related drugs is the second mechanism by which cross-sensitivity occurs. A classic example is hypersensitivity to aromatic antiepileptics (barbiturates, phenytoin, carbamazepine) with up to 50% reaction to a second drug in patients who reacted to one. For other drugs, in vitro and in vivo

data have suggested that cross-reactivity exists only between compounds with very similar chemical structures. Sulfamethoxazole-specific lymphocytes may be activated by other antibacterial sulfonamides but not diuretics, antidiabetic drugs, or anti-COX2 NSAIDs with a sulfonamide group. Approximately 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics.

Recent data suggest that although the risk of developing a drug eruption to another drug is increased in persons with a prior reaction, “cross-sensitivity” is probably not the explanation. As an example, those with a history of an allergic-like reaction to penicillin are at greater risk of developing a reaction to antibacterial sulfonamides than to cephalosporins.

These data suggest that the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLA genes, it is recommended that first-degree family members of patients with severe cutaneous reactions also should avoid causative agents. This may be most relevant for sulfonamides and antiepileptic medications.

ROLE OF TESTING FOR CAUSALITY AND DRUG RECHALLENGE

The usefulness of laboratory tests, skin-prick, or patch testing to determine causality is debated. Many in vitro immunologic assays have been developed for research purposes; however, the predictive value of these tests has not been validated in large series of affected patients. In some cases, diagnostic rechallenge may be appropriate, even for drugs with high rates of adverse reactions.

Skin-prick testing has clinical value in limited settings. In patients with a history suggesting immediate IgE-mediated reactions to penicillin, skin-prick testing with penicillins or cephalosporins has proven useful for identifying patients at risk of anaphylactic reactions to these agents. Negative skin tests do not totally rule out IgE-mediated reactivity; however, the risk of anaphylaxis in response to penicillin administration in patients with negative skin tests is about 1%. In contrast, two-thirds of patients with a positive skin test experience an allergic response upon rechallenge. The skin tests themselves carry a small risk of anaphylaxis.

For patients with delayed-type hypersensitivity, the clinical utility of skin tests remains questionable. At least one of a combination of several tests (prick, patch, and intradermal) is positive in 50–70% of patients with a reaction “definitely” attributed to a single medication. This low sensitivity corresponds to the observation that readministration of drugs with negative skin testing results in eruptions in 17% of cases.

Desensitization can be considered in those with a history of reaction to a medication that must be used again. Efficacy of such procedures has been demonstrated in cases of immediate reaction to penicillin and positive skin tests, anaphylactic reactions to platinum chemotherapy, and delayed reactions to sulfonamides in patients with AIDS. Desensitization is often successful in HIV-infected patients with morbilliform eruptions to sulfonamides but is not recommended in HIV-infected patients who developed erythroderma or a bullous reaction in response to prior sulfonamide exposure. Various protocols are available, including oral and

parenteral approaches. Oral desensitization appears to have a lower risk of serious anaphylactic reaction. Desensitization carries the risk of anaphylaxis regardless of how it is performed and should be performed in monitored clinical settings such as an intensive care unit. After desensitization, many patients experience non-life-threatening reactions during therapy with the culprit drug.

REPORTING

Any severe reaction to drugs should be reported to a regulatory agency or to pharmaceutical companies. Because severe reactions are too rare to be detected in premarketing clinical trials, spontaneous reports are of critical importance for early detection of unexpected life-threatening events. To be useful, the report should contain enough details to permit ascertainment of severity and drug causality.

ACKNOWLEDGMENT

We acknowledge the contribution of Drs. Jean-Claude Roujeau and Robert S. Stern to this chapter in previous editions.

FURTHER READING

Belum VR: Characterisation and management of dermatologic adverse events to agents targeting the PD-1 receptor. *Eur J Cancer* 60:12, 2016. [[PubMed: 27043866](#)]

Cornejo-Garcia JA et al: The genetics of drug hypersensitivity reactions. *J Investig Allergol Clin Immunol* 26:222, 2016. [[PubMed: 27375032](#)]

Creamer D et al: U.K. guidelines for the management of Stevens-Johnson syndrome/toxic epidermal necrolysis in adults 2016. *Br J Dermatol* 174:1194, 2016. [[PubMed: 27317286](#)]

Harp JL et al: Severe cutaneous adverse reactions: impact of immunology, genetics, and pharmacology. *Semin Cutan Med Surg* 33:17, 2014. [[PubMed: 25037255](#)]

Ko TM et al: Use of HLA-B*5801 genotyping to prevent [allopurinol](#) induced severe cutaneous adverse reactions in Taiwan: National prospective cohort study. *BMJ* 351:h4848, 2015. [[PubMed: 26399967](#)]

Lacouture ME et al: [Ipilimumab](#) in patients with cancer and the management of dermatologic adverse events. *J Am Acad Dermatol* 71:161, 2014. [[PubMed: 24767731](#)]

Mayorga C et al: In vitro tests for drug hypersensitivity reactions: An ENDA/EAACI Drug Allergy Interest Group position paper. *Allergy* 71:1103, 2016. [[PubMed: 26991315](#)]

Oussalah A et al: Genetic variants associated with drug-induced immediate hypersensitivity reactions: A PRISMA-compliant systematic review. Allergy 71:443, 2016. [[PubMed: 26678823](#)]

Petrelli F et al: Antibiotic prophylaxis for skin toxicity induced by antiepidermal growth factor receptor agents: A systematic review and meta-analysis. Br J Dermatol, 2016 ePub ahead of print. Accessed September 28, 2016.

Sassolas B et al: ALDEN, an algorithm for assessment of drug causality in Stevens-Johnson syndrome and toxic epidermal necrolysis: comparison with case-control analysis. Clin Pharmacol Ther 88:60, 2010. [[PubMed: 20375998](#)]

White KD et al: Evolving models of the immunopathogenesis of T cell-mediated drug allergy: The role of host, pathogens, and drug response. J Allergy Clin Immunol 136:219, 2015. [[PubMed: 26254049](#)]

Wolverton SE: Practice gaps: drug reactions. Dermatol Clin 34:311, 2016. [[PubMed: 27363888](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 63: Principles of Clinical Pharmacology

Dan M. Roden

INTRODUCTION

Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among healthcare providers and the lay community that the outcome of drug therapy varies widely among individuals. While this variability has been perceived as an unpredictable, and therefore inevitable, accompaniment of drug therapy, this is not the case. The goal of this chapter is to describe the principles of clinical pharmacology that can be used for the safe and optimal use of available and new drugs.


Drugs interact with specific target molecules to produce their beneficial and adverse effects. The chain of events between administration of a drug and production of these effects in the body can be divided into two components, both of which contribute to variability in drug actions. The first component comprises the processes that determine drug delivery to, and removal from, molecular targets. The resulting description of the relationship between drug concentration and time is termed *pharmacokinetics*. The second component of variability in drug action comprises the processes that determine variability in drug actions despite equivalent drug delivery to effector drug sites. This description of the relationship between drug concentration and effect is termed *pharmacodynamics*. As discussed further below, pharmacodynamic variability can arise as a result of variability in function of the target molecule itself or of variability in the broad biologic context in which the drug-target interaction occurs to achieve drug effects.

Two important goals of clinical pharmacology are (1) to provide a description of conditions under which drug actions vary among human subjects; and (2) to determine mechanisms underlying this variability, with the goal of improving therapy with available drugs as well as pointing to mechanisms whose targeting by new drugs may be effective in the treatment of human disease. The drug development process is briefly described at the end of this chapter.

The first steps in the discipline of clinical pharmacology were empirical descriptions of the influence of disease on drug actions and of individuals or families with unusual sensitivities to adverse drug effects. These important descriptive findings are now being replaced by an understanding of the molecular mechanisms underlying variability in drug actions. Importantly, it is often the personal interaction of the patient with the physician or other health care provider that first identifies unusual variability in drug actions; maintained alertness to unusual drug responses continues to be a key component of improving drug safety.

One useful unifying framework is to consider that the effects of disease, drug coadministration, or familial factors in modulating drug action reflect variability in expression or function of specific genes whose products determine pharmacokinetics and pharmacodynamics. This idea forms the basis for pharmacogenomic science; a few examples are cited in this chapter, and further details are addressed in [Chap. 64](#).

GLOBAL CONSIDERATIONS

 It is true across all cultures and diseases that factors such as compliance, genetic variants affecting pharmacokinetics, or pharmacodynamics (which themselves vary by ancestry), and drug interactions contribute to drug responses. Cost issues or cultural factors may determine the likelihood that specific drugs, drug combinations, or over-the-counter (OTC) remedies are prescribed. The broad principles of clinical pharmacology enunciated here can be used to analyze the mechanisms underlying successful or unsuccessful therapy with any drug.

INDICATIONS FOR DRUG THERAPY: RISK VERSUS BENEFIT

It is self-evident that the benefits of drug therapy should outweigh the risks. Benefits fall into two broad categories: those designed to alleviate a symptom and those designed to prolong useful life. An increasing emphasis on the principles of evidence-based medicine and techniques such as large clinical trials and meta-analyses has defined benefits of drug therapy in broad patient populations. However, establishing the balance between risk and benefit is not always simple. An increasing body of evidence supports the idea, with which practitioners are very familiar, that individual patients may display responses that are not expected from large population studies and often have comorbidities that typically exclude them from large clinical trials. In addition, therapies that provide symptomatic benefits but shorten life may be entertained in patients with serious and highly symptomatic diseases such as heart failure or cancer. These considerations illustrate the continuing, highly personal nature of the relationship between the prescriber and the patient.

Adverse Effects

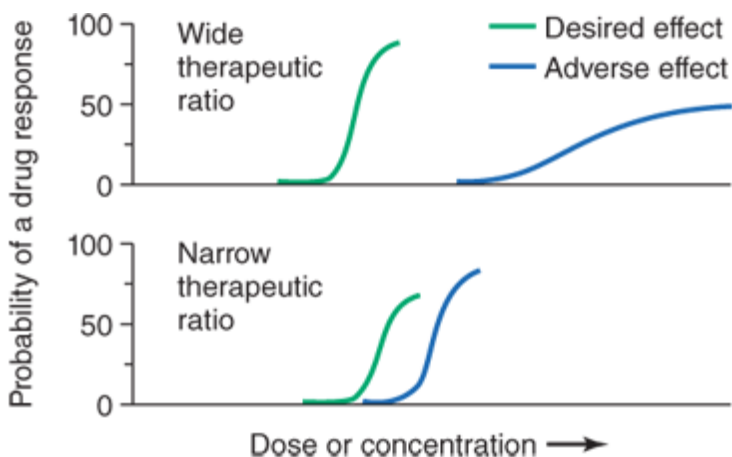
Some adverse effects are so common and so readily associated with drug therapy that they are identified very early during clinical use of a drug. By contrast, serious adverse drug reactions may be sufficiently uncommon that they escape detection for many years after a drug begins to be widely used. The issue of how to identify rare but serious adverse effects (that can profoundly affect the benefit-risk perception in an individual patient) has not been satisfactorily resolved. Potential approaches range from an increased understanding of the molecular and genetic basis of variability in drug actions to expanded post-marketing surveillance mechanisms. None of these have been completely effective, so practitioners must be continuously vigilant to the possibility that unusual symptoms may be related to specific drugs, or combinations of drugs, that their patients receive.

Therapeutic Index

Beneficial and adverse reactions to drug therapy can be described by a series of dose-response relations (Fig. 63-1). Well-tolerated drugs demonstrate a wide margin, termed the *therapeutic ratio*, *therapeutic index*, or *therapeutic window*, between the doses required to produce a therapeutic effect and those producing toxicity. In cases where there is a similar relationship between plasma drug concentration and effects, monitoring plasma concentrations can be a highly effective aid in managing drug therapy by enabling concentrations to be maintained above the minimum required to produce an effect and below the concentration range likely to produce toxicity. Such monitoring has been widely used to guide therapy with specific agents, such as certain antiarrhythmics, anticonvulsants, and antibiotics. Many of the principles in clinical pharmacology and examples outlined below, which can be applied broadly to therapeutics, have been developed in these arenas.

FIGURE 63-1

The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. **Top.** A drug with a wide therapeutic ratio, that is, a wide separation of the two curves. **Bottom.** A drug with a narrow therapeutic ratio; here, the likelihood of adverse effects at therapeutic doses is increased because the curves are not well separated. Further, a steep dose-response curve for adverse effects is especially undesirable, as it implies that even small dosage increments may sharply increase the likelihood of toxicity. When there is a definable relationship between drug concentration (usually measured in plasma) and desirable and adverse effect curves, concentration may be substituted on the abscissa. Note that not all patients necessarily demonstrate a therapeutic response (or adverse effect) at any dose, and that some effects (notably some adverse effects) may occur in a dose-independent fashion.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

PRINCIPLES OF PHARMACOKINETICS

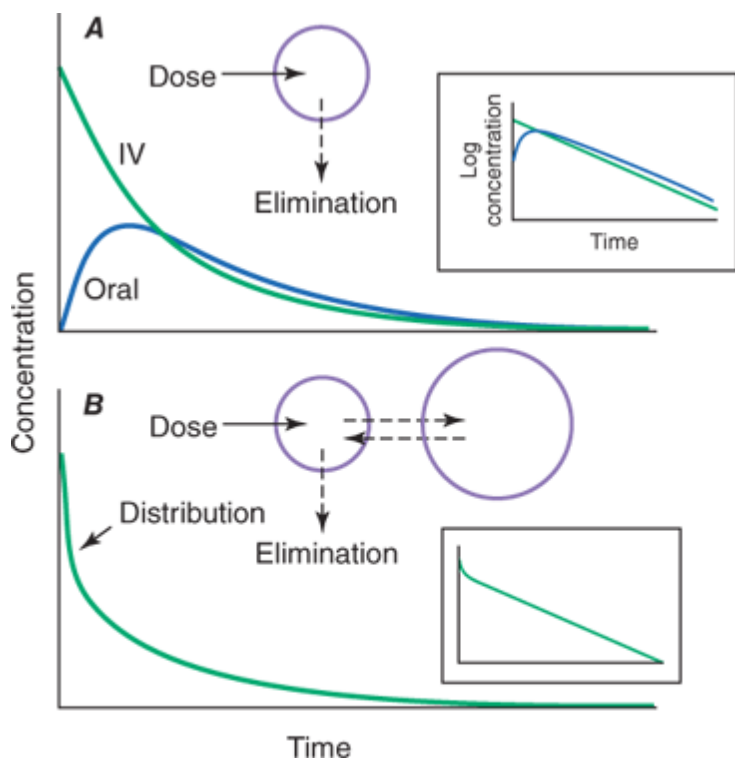
The processes of absorption, distribution, metabolism, and excretion—collectively termed *drug disposition*—determine the concentration of drug delivered to target effector molecules.

ABSORPTION AND BIOAVAILABILITY

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug actually entering the systemic circulation may be less than with the intravenous route (Fig. 63-2A). The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) absorption is reduced, or (2) the drug undergoes metabolism or elimination prior to entering the systemic circulation. Occasionally, the administered drug formulation is inconsistent or has degraded with time; for example, the anticoagulant dabigatran degrades rapidly (over weeks) once exposed to air, so the amount administered may be less than prescribed.

FIGURE 63-2

Idealized time-plasma concentration curves after a single dose of drug. A. The time course of drug concentration after an instantaneous IV bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV, indicating incomplete bioavailability. Note that despite this incomplete bioavailability, concentration after the oral dose can be higher than after the IV dose at some time points. The inset shows that the decline of concentrations over time is linear on a log-linear plot, characteristic of first-order elimination, and that oral and IV drugs have the same elimination (parallel) time course. **B.** The decline of central compartment concentration when drug is distributed both to and from a peripheral compartment and eliminated from the central compartment. The rapid initial decline of concentration reflects not drug elimination but distribution.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

When a drug is administered by a non-intravenous route, the peak concentration occurs later and is lower than after the same dose given by rapid intravenous injection, reflecting absorption from the site of administration (Fig. 63-2). The extent of absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at its site of administration, or has physicochemical properties

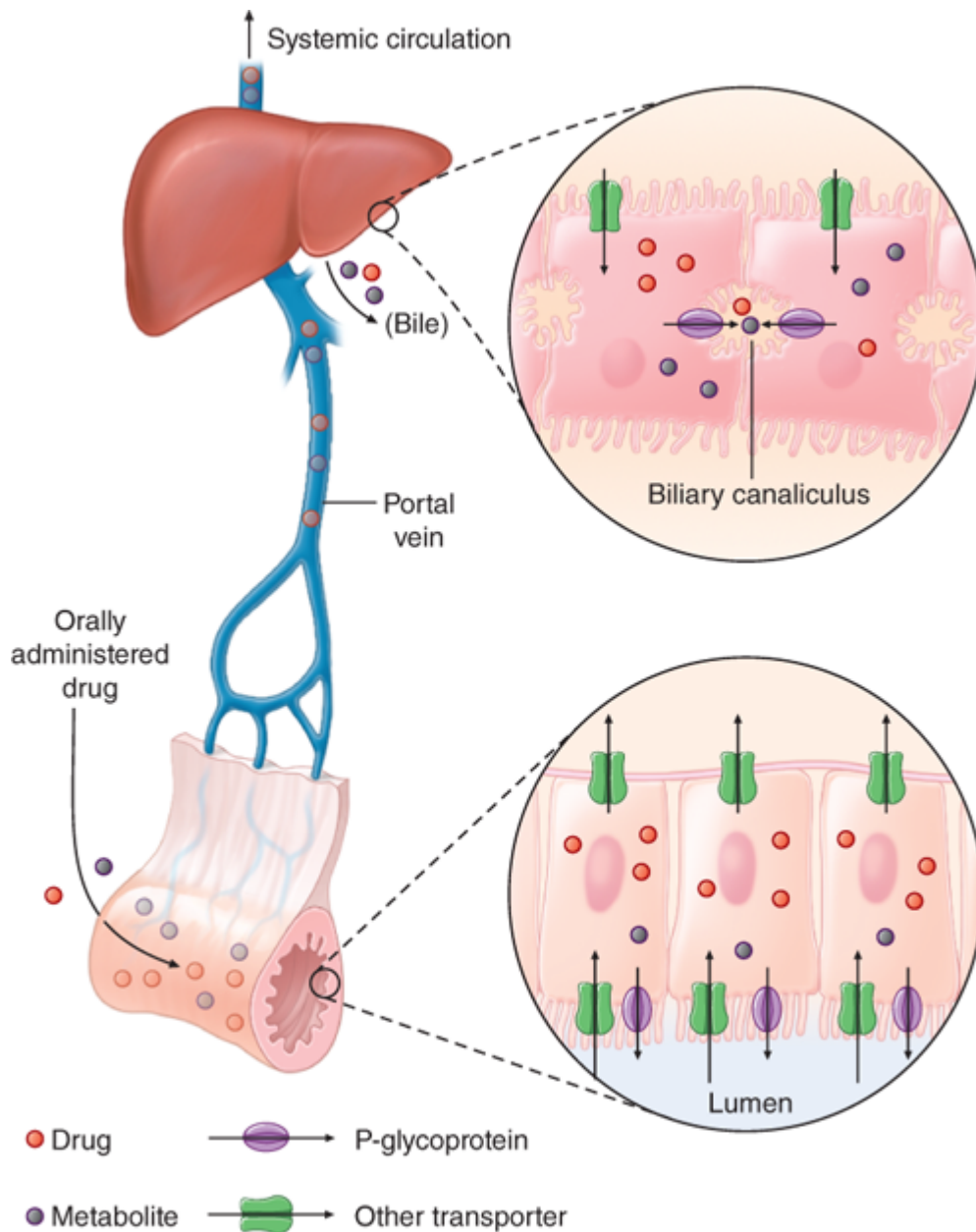
such as insolubility that prevent complete absorption from its site of administration. Slow absorption rates are deliberately designed into “slow-release” or “sustained-release” drug formulations in order to minimize variation in plasma concentrations during the interval between doses.

“First-Pass” Effect

When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the liver prior to entering the systemic circulation (**Fig. 63-3**). Once a drug enters the enterocyte, it may undergo metabolism, be transported into the portal vein, or be excreted back into the intestinal lumen. Both excretion into the intestinal lumen and metabolism decrease systemic bioavailability. Once a drug passes this enterocyte barrier, it may also be taken up into the hepatocyte, where bioavailability can be further limited by metabolism or excretion into the bile. This elimination in intestine and liver, which reduces the amount of drug delivered to the systemic circulation, is termed *presystemic elimination*, *presystemic extraction*, or *first-pass elimination*.

FIGURE 63-3

Mechanism of presystemic clearance. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to the systemic circulation. (*Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

DRUG TRANSPORT

Drug movement across the membrane of any cell, including enterocytes and hepatocytes, is a combination of passive diffusion and active transport, mediated by specific drug uptake and efflux molecules. One widely studied drug transport molecule is the drug efflux pump P-glycoprotein, the product of the *ABCB1* (or *MDR1*) gene. P-glycoprotein is expressed on the apical aspect of the enterocyte and on the canalicular aspect of the hepatocyte (Fig. 63-3). In both locations, it serves as an efflux pump, limiting availability of drug to the systemic circulation. P-glycoprotein-mediated drug efflux from cerebral capillaries limits drug brain penetration and is an important component of the blood-brain barrier. Other transporters mediate uptake into cells of drugs and endogenous substrates such as vitamins or nutrients.

DRUG METABOLISM

Drug metabolism generates compounds that are usually more polar and, hence, more readily excreted than parent drug. Metabolism takes place predominantly in the liver but can occur at other sites such as kidney, intestinal epithelium, lung, and plasma. “Phase I” metabolism involves chemical modification, most often oxidation accomplished by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs and other molecules that are especially important for drug metabolism are presented in [Table 63-1](#), and each drug may be a substrate for one or more of these enzymes. “Phase II” metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites may exert important pharmacologic activity, as discussed further below.

TABLE 63-1

Molecular Pathways Mediating Drug Disposition

MOLECULE	SUBSTRATES ^a	INHIBITORS ^a
CYP3A	Calcium channel blockers	Amiodarone
	Antiarrhythmics (lidocaine , quinidine, mexiletine)	Ketoconazole , itraconazole
	HMG-CoA reductase inhibitors (“statins”; see text)	Erythromycin , clarithromycin
	Cyclosporine , tacrolimus	Ritonavir
	Indinavir, saquinavir , ritonavir	
CYP2D6 ^b	Timolol , metoprolol , carvedilol Propafenone, flecainide Tricyclic antidepressants Fluoxetine, paroxetine	Quinidine (even at ultra-low doses) Tricyclic antidepressants Fluoxetine, paroxetine
CYP2C9 ^b	Warfarin	Amiodarone
	Phenytoin	Fluconazole
	Glipizide	Phenytoin
	Losartan	
CYP2C19 ^b	Omeprazole	Omeprazole
	Mephenytoin	
	Clopidogrel	
CYP2B6 ^b	Efavirenz	
Thiopurine S-methyltransferase ^b	6-Mercaptopurine, azathioprine	

MOLECULE	SUBSTRATES ^a	INHIBITORS ^a
N-acetyltransferase ^b	Isoniazid	
	Procainamide	
	Hydralazine	
	Some sulfonamides	
UGT1A1 ^b	Irinotecan	
Pseudocholinesterase ^b	Succinylcholine	
P-glycoprotein	Digoxin	Quinidine
	HIV protease inhibitors	Amiodarone
	Many CYP3A substrates	Verapamil
		Cyclosporine
		Itraconazole
		Erythromycin
SLCO1B1 ^b	Simvastatin and some other statins	

^aInhibitors affect the molecular pathway, and thus may affect substrate. ^bClinically important genetic variants described; see **Chap. 64**.

Note: A listing of CYP substrates, inhibitors, and inducers is maintained at <http://medicine.iupui.edu/clinpharm/ddis/main-table>.

Clinical Implications of Altered Bioavailability

Some drugs undergo near-complete presystemic metabolism and thus cannot be administered orally. **Nitroglycerin** cannot be used orally because it is completely extracted prior to reaching the systemic circulation. The drug is, therefore, used by the sublingual, transdermal, or intravascular routes, which bypass presystemic metabolism.

Some drugs with very extensive presystemic metabolism can still be administered by the oral route, using much higher doses than those required intravenously. Thus, a typical intravenous dose of [verapamil](#) is 1–5 mg, compared to a usual single oral dose of 40–120 mg. Administration of low-dose aspirin can result in exposure of cyclooxygenase in platelets in the portal vein to the drug, but systemic sparing because of first-pass aspirin deacylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

HALF-LIFE

Most pharmacokinetic processes, such as elimination, are first-order; that is, the rate of the process depends on the amount of drug present. Elimination can occasionally be zero-order (fixed amount eliminated per unit time), and this can be clinically important (see “Principles of Dose Selection”). In the simplest pharmacokinetic model ([Fig. 63-2A](#)), a drug bolus (D) is administered instantaneously to a central compartment, from which drug elimination occurs as a first-order process. Occasionally, central and other compartments correspond to physiologic spaces (e.g., plasma volume), whereas in other cases they are simply mathematical functions used to describe drug disposition. The first-order nature of drug elimination leads directly to the relationship describing drug concentration (C) at any time (t) following the bolus:

$$C = \frac{D}{V_c} \cdot e^{(-0.69t/t_{1/2})}$$

where V_c is the volume of the compartment into which drug is delivered and $t_{1/2}$ is elimination half-life. As a consequence of this relationship, a plot of the logarithm of concentration versus time is a straight line ([Fig. 63-2A, inset](#)). *Half-life* is the time required for 50% of a first-order process to be completed. Thus, 50% of drug elimination is achieved after one drug-elimination half-life, 75% after two, 87.5% after three, etc. In practice, first-order processes such as elimination are near-complete after four–five half-lives.

In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this case, the plot of plasma concentration versus time after a bolus may demonstrate two (or more) exponential components ([Fig. 63-2B](#)). In general, the initial rapid drop in drug concentration represents not elimination but drug distribution into and out of peripheral tissues (also first-order processes), while the slower component represents drug elimination; the initial precipitous decline is usually evident with administration by intravenous but not by other routes. Drug concentrations at peripheral sites are determined by a balance between drug distribution to and redistribution from those sites, as well as by elimination. Once distribution is near-complete (four–five distribution half-lives), plasma and tissue concentrations decline in parallel.

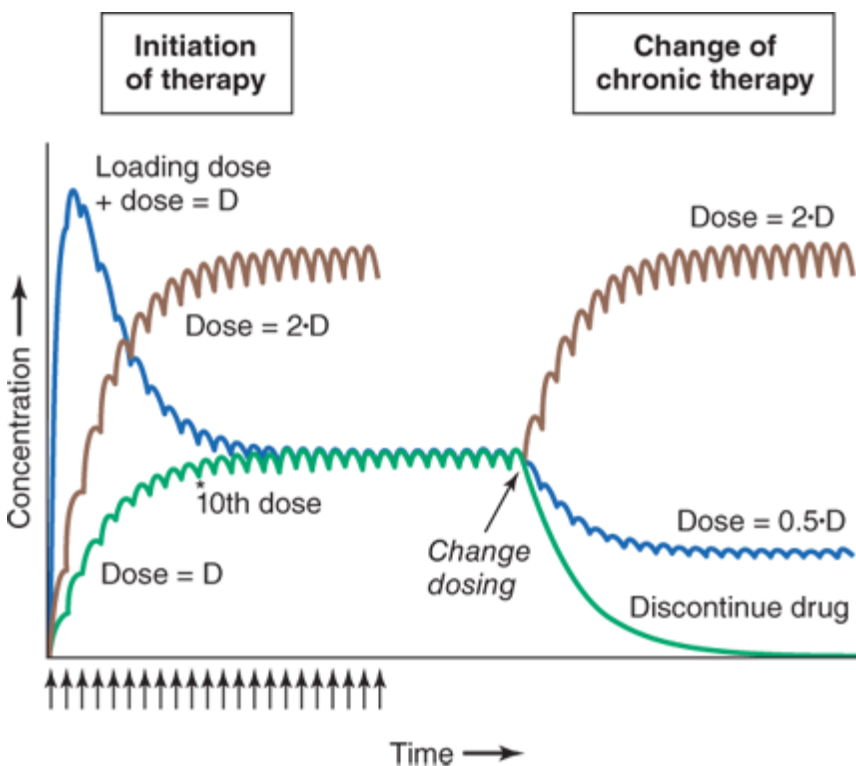
Clinical Implications of Half-Life Measurements

The elimination half-life not only determines the time required for drug concentrations to fall to near-immeasurable levels after a single bolus, it is also the sole determinant of the time required for steady-state plasma concentrations to be achieved after any change in drug dosing ([Fig. 63-4](#)). This applies to the

initiation of chronic drug therapy (whether by multiple oral doses or by continuous intravenous infusion), a change in chronic drug dose or dosing interval, or discontinuation of drug.

FIGURE 63-4

Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state achieved. A doubling of the dose resulted in a doubling of the steady state but the same time course of accumulation. Once steady state is achieved, a change in dose (increase, decrease, or drug discontinuation) results in a new steady state in ~5 elimination half-lives. (Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: *Cardiac Electrophysiology: From Cell to Bedside*, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: *Harrison's Principles of Internal Medicine*, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Steady state describes the situation during chronic drug administration when the amount of drug administered per unit time equals drug eliminated per unit time. With a continuous intravenous infusion, plasma concentrations at steady state are stable, while with chronic oral drug administration, plasma concentrations vary during the dosing interval but the time-concentration profile between dosing intervals is stable (Fig. 63-4).

DRUG DISTRIBUTION

In a typical 70-kg human, plasma volume is ~3 L, blood volume is ~5.5 L, and extracellular water outside the vasculature is ~20 L. The volume of distribution of drugs extensively bound to plasma proteins but not to tissue components approaches plasma volume; warfarin is an example. By contrast, for drugs highly bound to tissues, the volume of distribution can be far greater than any physiologic space. For example, the volume

of distribution of digoxin and tricyclic antidepressants is hundreds of liters, obviously exceeding total-body volume. Such drugs are not readily removed by dialysis, an important consideration in overdose.

Clinical Implications of Drug Distribution

In some cases, pharmacologic effects require drug distribution to peripheral sites. In this instance, the time course of drug delivery to and removal from these sites determines the time course of drug effects; anesthetic uptake into the central nervous system (CNS) is an example.

LOADING DOSES

For some drugs, the indication may be so urgent that administration of “loading” dosages is required to achieve rapid elevations of drug concentration and therapeutic effects earlier than with chronic maintenance therapy (Fig. 63-4). Nevertheless, the time required for true steady state to be achieved is still determined only by the elimination half-life.

RATE OF INTRAVENOUS DRUG ADMINISTRATION

Although the simulations in Fig. 63-2 use a single intravenous bolus, this is usually inappropriate in practice because side effects related to transiently very high concentrations can result. Rather, drugs are more usually administered orally or as a slower intravenous infusion. Some drugs are so predictably lethal when infused too rapidly that special precautions should be taken to prevent accidental boluses. For example, solutions of potassium for intravenous administration >20 mEq/L should be avoided in all but the most exceptional and carefully monitored circumstances. This minimizes the possibility of cardiac arrest due to accidental increases in infusion rates of more concentrated solutions.

Transiently high drug concentrations after rapid intravenous administration can occasionally be used to advantage. The use of midazolam for intravenous sedation, for example, depends upon its rapid uptake by the brain during the distribution phase to produce sedation quickly, with subsequent egress from the brain during the redistribution of the drug as equilibrium is achieved.

Similarly, adenosine must be administered as a rapid bolus in the treatment of reentrant supraventricular tachycardias (Chap. 241) to prevent elimination by very rapid ($t_{1/2}$ of seconds) uptake into erythrocytes and endothelial cells before the drug can reach its clinical site of action, the atrioventricular node.

Clinical Implications of Altered Protein Binding

Many drugs circulate in the plasma partly bound to plasma proteins. Since only unbound (free) drug can distribute to sites of pharmacologic action, drug response is related to the free rather than the total circulating plasma drug concentration. In chronic kidney or liver disease, protein binding may be decreased and thus drug actions increased. In some situations (myocardial infarction, infection, surgery), acute phase reactants transiently increase binding of some drugs and thus decrease efficacy. These changes assume the greatest clinical importance for drugs that are highly protein-bound since even a small change in protein binding can result in large changes in free drug; for example, a decrease in binding from 99 to 98% doubles

the free drug concentration from 1 to 2%. For some drugs (e.g., phenytoin), monitoring free rather than total drug concentrations can be useful.

DRUG ELIMINATION

Drug elimination reduces the amount of drug in the body over time. An important approach to quantifying this reduction is to consider that drug concentrations at the beginning and end of a time period are unchanged and that a specific volume of the body has been “cleared” of the drug during that time period. This defines clearance as volume/time. Clearance includes both drug metabolism and excretion.

Clinical Implications of Altered Clearance

While elimination half-life determines the time required to achieve steady-state plasma concentration (C_{ss}), the *magnitude* of that steady state is determined by clearance (Cl) and dose alone. For a drug administered as an intravenous infusion, this relationship is:

$$C_{ss} = \text{dosing rate} / Cl \text{ or } \text{dosing rate} = Cl \cdot C_{ss}$$

When drug is administered orally, the average plasma concentration within a dosing interval ($C_{avg,ss}$) replaces C_{ss} , and the dosage (dose per unit time) must be increased if bioavailability (F) is <100%:

$$\text{Dose/time} = Cl \cdot C_{avg,ss} / F$$

Genetic variants, drug interactions, or diseases that reduce the activity of drug-metabolizing enzymes or excretory mechanisms lead to decreased clearance and, hence, a requirement for downward dose adjustment to avoid toxicity. Conversely, some drug interactions and genetic variants increase the function of drug elimination pathways, and hence, increased drug dosage is necessary to maintain a therapeutic effect.

ACTIVE DRUG METABOLITES

Metabolites may produce effects similar to, overlapping with, or distinct from those of the parent drug. Accumulation of the major metabolite of procainamide, *N*-acetylprocainamide (NAPA), likely accounts for marked QT prolongation and torsades des pointes ventricular tachycardia ([Chap. 247](#)) during therapy with procainamide. Neurotoxicity during therapy with the opioid analgesic meperidine is likely due to accumulation of normeperidine, especially in renal disease.

Prodrugs are inactive compounds that require metabolism to generate active metabolites that mediate the drug effects. Examples include many angiotensin-converting enzyme (ACE) inhibitors, the angiotensin receptor blocker losartan, the antineoplastic irinotecan, the anti-estrogen tamoxifen, the analgesic codeine

(whose active metabolite morphine probably underlies the opioid effect during codeine administration), and the antiplatelet drug clopidogrel. Drug metabolism has also been implicated in bioactivation of procarcinogens and in generation of reactive metabolites that mediate certain adverse drug effects (e.g., acetaminophen hepatotoxicity, discussed below).

THE CONCEPT OF HIGH-RISK PHARMACOKINETICS

When plasma concentrations of active drug depend exclusively on a single metabolic pathway, any condition that inhibits that pathway (be it disease-related, genetic, or due to a drug interaction) can lead to dramatic changes in drug concentrations and marked variability in drug action. Two mechanisms can generate highly variable drug concentrations and effects through such “high-risk pharmacokinetics.” *First*, variability in bioactivation of a prodrug can lead to striking variability in drug action; examples include decreased CYP2D6 activity, which prevents analgesia by codeine, and decreased CYP2C19 activity, which reduces the antiplatelet effects of clopidogrel. The *second* setting is drug elimination that relies on a single pathway. In this case, inhibition of the elimination pathway by genetic variants or by administration of inhibiting drugs leads to marked elevation of drug concentration and, for drugs with a narrow therapeutic window, an increased likelihood of dose-related toxicity. The active S-enantiomer of the anticoagulant warfarin is eliminated by CYP2C9, and co-administration of amiodarone or phenytoin, CYP2C9 inhibitors, may therefore increase the risk of bleeding unless the dose is decreased. When drugs undergo elimination by multiple-drug metabolizing or excretory pathways, absence of one pathway (due to a genetic variant or drug interaction) is much less likely to have a large impact on drug concentrations or drug actions.

PRINCIPLES OF PHARMACODYNAMICS

The Onset of Drug Action

For drugs used in the urgent treatment of acute symptoms, little or no delay is anticipated (or desired) between the drug-target interaction and the development of a clinical effect. Examples of such acute situations include vascular thrombosis, shock, or status epilepticus.

For many conditions, however, the indication for therapy is less urgent, and a delay between the interaction of a drug with its pharmacologic target(s) and a clinical effect is clinically acceptable. Common pharmacokinetic mechanisms that can contribute to such a delay include slow elimination (resulting in slow accumulation to steady state), uptake into peripheral compartments, or accumulation of active metabolites. A common pharmacodynamic explanation for such a delay is that the clinical effect develops as a downstream consequence of the initial molecular effect the drug produces. Thus, administration of a proton pump inhibitor or an H₂-receptor blocker produces an immediate increase in gastric pH but ulcer healing that is delayed. Cancer chemotherapy similarly produces delayed therapeutic effects.

Drug Effects May Be Disease Specific

A drug may produce no action or a different spectrum of actions in unaffected individuals compared to patients with underlying disease. Further, concomitant disease can complicate interpretation of response to drug therapy, especially adverse effects. For example, high doses of anticonvulsants such as phenytoin may cause neurologic symptoms, which may be confused with the underlying neurologic disease. Similarly, increasing dyspnea in a patient with chronic lung disease receiving amiodarone therapy could be due to the drug, underlying disease, or an intercurrent cardiopulmonary problem. As a result, alternate antiarrhythmic therapies are preferable in patients with chronic lung disease.

While drugs interact with specific molecular receptors, drug effects may vary over time, even if stable drug and metabolite concentrations are maintained. The drug-receptor interaction occurs in a complex biologic milieu that can vary to modulate the drug effect. For example, ion channel blockade by drugs, an important anticonvulsant and antiarrhythmic effect, is often modulated by membrane potential, itself a function of factors such as extracellular potassium or local ischemia. Receptors may be up- or down-regulated by disease or by the drug itself. For example, β -adrenergic blockers upregulate β -receptor density during chronic therapy. While this effect does not usually result in resistance to the therapeutic effect of the drugs, it may produce severe agonist-mediated effects (such as hypertension or tachycardia) if the blocking drug is abruptly withdrawn.

PRINCIPLES OF DOSE SELECTION

The desired goal of therapy with any drug is to maximize the likelihood of a beneficial effect while minimizing the risk of adverse effects. Previous experience with the drug, in controlled clinical trials or in post-marketing use, defines the relationships between dose or plasma concentration and these dual effects (Fig. 63-1) and has important implications for initiation of drug therapy:

1. *The target drug effect should be defined when drug treatment is started.* With some drugs, the desired effect may be difficult to measure objectively, or the onset of efficacy can be delayed for weeks or months; drugs used in the treatment of cancer and psychiatric disease are examples. Sometimes a drug is used to treat a symptom, such as pain or palpitations, and here it is the patient who will report whether the selected dose is effective. In yet other settings, such as anticoagulation or hypertension, the desired response can be repeatedly and objectively assessed by simple clinical or laboratory tests.
2. *The nature of anticipated toxicity often dictates the starting dose.* If side effects are minor, it may be acceptable to start chronic therapy at a dose highly likely to achieve efficacy and down-titrate if side effects occur. However, this approach is rarely, if ever, justified if the anticipated toxicity is serious or life-threatening; in this circumstance, it is more appropriate to initiate therapy with the lowest dose that may produce a desired effect. In cancer chemotherapy, it is common practice to use maximum-tolerated doses.
3. *The above considerations do not apply if these relationships between dose and effects cannot be defined.* This is especially relevant to some adverse drug effects (discussed further below) whose development is not readily related to drug dose.

4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small.*

Failure of Efficacy

Assuming the diagnosis is correct and the correct drug is prescribed, explanations for failure of efficacy include drug interactions, noncompliance, or unexpectedly low drug concentration due to administration of expired or degraded drug. These are situations in which measurement of plasma drug concentrations, if available, can be especially useful. Noncompliance is an especially frequent problem in the long-term treatment of diseases such as hypertension and epilepsy, occurring in $\geq 25\%$ of patients in therapeutic environments in which no special effort is made to involve patients in the responsibility for their own health. Multidrug regimens with multiple doses per day are especially prone to noncompliance.

Monitoring response to therapy, by physiologic measures or by plasma concentration measurements, requires an understanding of the relationships between plasma concentration and anticipated effects. For example, measurement of QT interval is used during treatment with sotalol or dofetilide to avoid marked QT prolongation that can herald serious arrhythmias. In this setting, evaluating the electrocardiogram at the time of anticipated peak plasma concentration and effect (e.g., 1–2 h post-dose at steady state) is most appropriate. Maintained high vancomycin levels carry a risk of nephrotoxicity, so dosages should be adjusted on the basis of plasma concentrations measured at trough (pre-dose). Similarly, for dose adjustment of other drugs (e.g., anticonvulsants), concentration should be measured at its lowest during the dosing interval, just prior to a dose at steady state (Fig. 63-4), to ensure a maintained therapeutic effect.

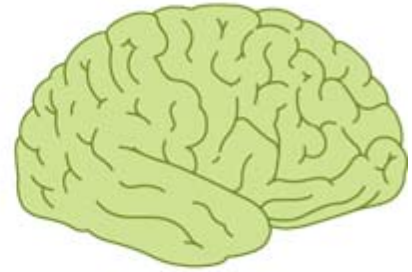
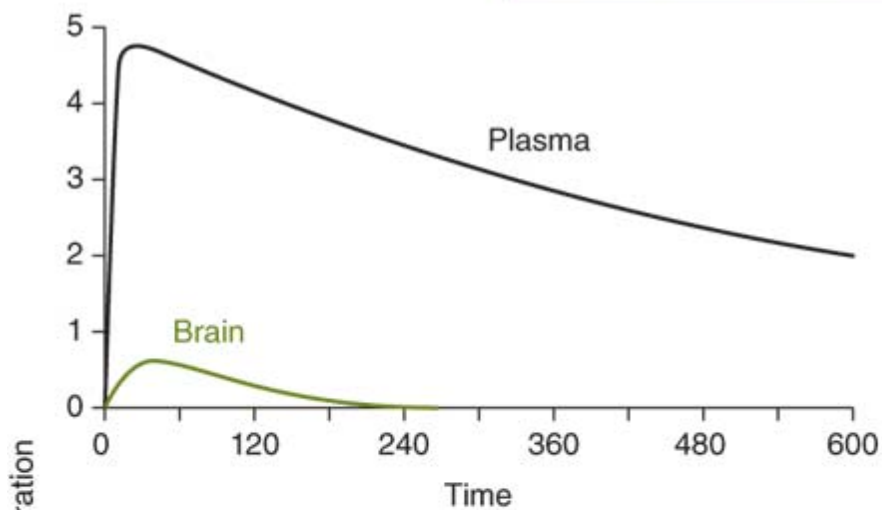
Concentration of Drugs in Plasma as a Guide to Therapy

Factors such as interactions with other drugs, disease-induced alterations in elimination and distribution, and genetic variation in drug disposition combine to yield a wide range of plasma levels in patients given the same dose. Hence, if a predictable relationship can be established between plasma drug concentration and beneficial or adverse drug effect, measurement of plasma levels can provide a valuable tool to guide selection of an optimal dose, especially when there is a narrow range between the plasma levels yielding therapeutic and adverse effects. Monitoring is commonly used with certain types of drugs including many anticonvulsants, antirejection agents, antiarrhythmics, and antibiotics. By contrast, if no such relationship can be established (e.g., if drug access to important sites of action outside plasma is highly variable), monitoring plasma concentration may not provide an accurate guide to therapy (Fig. 63-5).

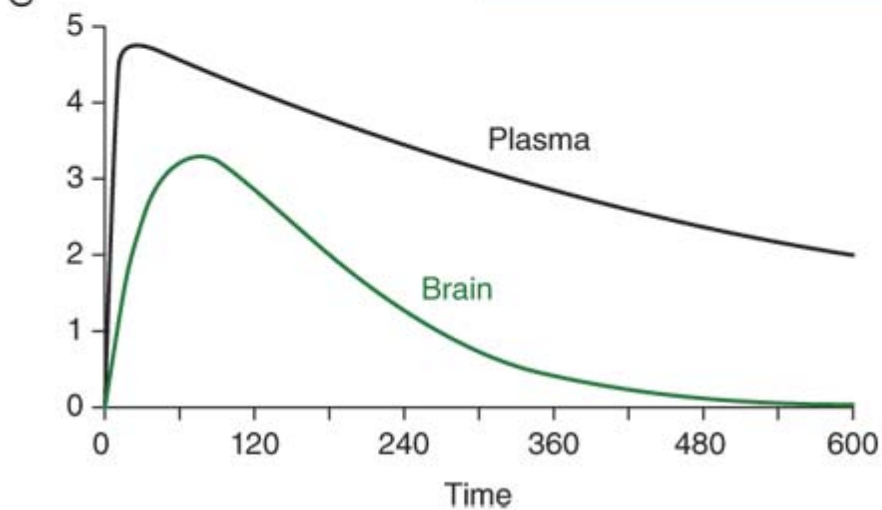
FIGURE 63-5

The efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain and so constitutes a key element of the blood-brain barrier. Thus, reduced P-glycoprotein function (e.g., due to drug interactions) increases penetration of substrate drugs into the brain, even when plasma concentrations are unchanged.

Normal P-glycoprotein function



Decreased P-glycoprotein function



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the dosing rate. Accordingly, the maintenance dose may be adjusted on the basis of the ratio between the desired and measured concentrations *at steady state*; for example, if a doubling of the steady-state plasma concentration is desired, the dose should be doubled. This does not apply to drugs eliminated by zero-order kinetics (fixed amount per unit time), where small dosage increases will produce disproportionate increases in plasma concentration; examples include phenytoin and theophylline.

An increase in dosage is usually best achieved by changing the drug dose but not the dosing interval (e.g., by giving 200 mg every 8 h instead of 100 mg every 8 h). However, this approach is acceptable only if the resulting maximum concentration is not toxic and the trough value does not fall below the minimum effective concentration for an undesirable period of time. Alternatively, the steady state may be changed by altering the frequency of intermittent dosing but not the size of each dose. In this case, the magnitude of the

fluctuations around the average steady-state level will change—the shorter the dosing interval, the smaller the difference between peak and trough levels.

EFFECTS OF DISEASE ON DRUG CONCENTRATION AND RESPONSE

RENAL DISEASE

Renal excretion of parent drug and metabolites is generally accomplished by glomerular filtration and by specific drug transporters. If a drug or its metabolites are primarily excreted through the kidneys and increased drug levels are associated with adverse effects (an example of “high-risk pharmacokinetics” described above), drug dosages must be reduced in patients with renal dysfunction to avoid toxicity. The antiarrhythmics dofetilide and sotalol undergo predominant renal excretion and carry a risk of QT prolongation and arrhythmias if doses are not reduced in renal disease. In end-stage renal disease, sotalol has been given as 40 mg after dialysis (every second day), compared to the usual daily dose, 80–120 mg every 12 h. At approved doses, the anticoagulant edoxaban appears to be somewhat more effective in subjects with mild renal dysfunction, possibly reflecting higher drug levels. The narcotic analgesic meperidine undergoes extensive hepatic metabolism, so that renal failure has little effect on its plasma concentration. However, its metabolite, normeperidine, does undergo renal excretion, accumulates in renal failure, and probably accounts for the signs of CNS excitation, such as irritability, twitching, and seizures, that appear when multiple doses of meperidine are administered to patients with renal disease. Protein binding of some drugs (e.g., phenytoin) may be altered in uremia, so measuring free drug concentration may be desirable.

In non-end-stage renal disease, changes in renal drug clearance are generally proportional to those in creatinine clearance, which may be measured directly or estimated from the serum creatinine. This estimate, coupled with the knowledge of how much drug is normally excreted renally versus non-renally, allows an estimate of the dose adjustment required. In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance. Any such modification of dose is a first approximation and should be followed by plasma concentration data (if available) and clinical observation to further optimize therapy for the individual patient.

LIVER DISEASE

Standard tests of liver function are not useful in adjusting doses in diseases like hepatitis or cirrhosis. First-pass metabolism may decrease, leading to increased oral bioavailability as a consequence of disrupted hepatocyte function, altered liver architecture, and portacaval shunts. The oral bioavailability for high first-pass drugs such as morphine, meperidine, midazolam, and nifedipine is almost doubled in patients with cirrhosis, compared to those with normal liver function. Therefore, the size of the oral dose of such drugs should be reduced in this setting.

HEART FAILURE AND SHOCK

Under conditions of decreased tissue perfusion, the cardiac output is redistributed to preserve blood flow to the heart and brain at the expense of other tissues ([Chap. 252](#)). As a result, drugs may be distributed into a smaller volume of distribution, higher drug concentrations will be present in the plasma, and the tissues that are best perfused (the brain and heart) will be exposed to these higher concentrations, resulting in increased CNS or cardiac effects. As well, decreased perfusion of the kidney and liver may impair drug clearance. Another consequence of severe heart failure is decreased gut perfusion, which may reduce drug absorption and, thus, lead to reduced or absent effects of orally administered therapies.

DRUG USE IN THE ELDERLY

In the elderly, multiple pathologies and medications used to treat them result in more drug interactions and adverse effects. Aging also results in changes in organ function, especially of the organs involved in drug disposition. Initial doses should be less than the usual adult dosage and should be increased slowly. The number of medications, and doses per day, should be kept as low as possible.

Even in the absence of kidney disease, renal clearance may be reduced by 35–50% in elderly patients. Dosages should be adjusted on the basis of creatinine clearance. Aging also results in a decrease in the size of, and blood flow to, the liver and possibly in the activity of hepatic drug-metabolizing enzymes; accordingly, the hepatic clearance of some drugs is impaired in the elderly. As with liver disease, these changes are not readily predicted.

Elderly patients may display altered drug sensitivity. Examples include increased analgesic effects of opioids, increased sedation from benzodiazepines and other CNS depressants, and increased risk of bleeding while receiving anticoagulant therapy, even when clotting parameters are well controlled. Exaggerated responses to cardiovascular drugs are also common because of the impaired responsiveness of normal homeostatic mechanisms. Conversely, the elderly display decreased sensitivity to β -adrenergic receptor blockers.

Adverse drug reactions are especially common in the elderly because of altered pharmacokinetics and pharmacodynamics, the frequent use of multidrug regimens, and concomitant disease. For example, use of long half-life benzodiazepines is linked to the occurrence of hip fractures in elderly patients, perhaps reflecting both a risk of falls from these drugs (due to increased sedation) and the increased incidence of osteoporosis in elderly patients. In population surveys of the noninstitutionalized elderly, as many as 10% had at least one adverse drug reaction in the previous year.

DRUG USE IN CHILDREN

While most drugs used to treat disease in children are the same as those in adults, there are few studies that provide solid data to guide dosing. Drug metabolism pathways mature at different rates after birth, and disease mechanisms may be different in children. In practice, doses are adjusted for size (weight or body surface area) as a first approximation unless age-specific data are available.

INTERACTIONS BETWEEN DRUGS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels.

Interactions must be considered in the differential diagnosis of any unusual response occurring during drug therapy. Prescribers should recognize that patients often come to them with a legacy of drugs acquired during previous medical experiences, often with multiple physicians who may not be aware of all the patient's medications. A meticulous drug history should include examination of the patient's medications and, if necessary, calls to the pharmacist to identify prescriptions. It should also address the use of agents not often volunteered during questioning, such as OTC drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to memorize these, certain drugs consistently run the risk of generating interactions, often by inhibiting or inducing specific drug elimination pathways. Examples are presented below and in **Table 63-2**. Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

TABLE 63-2

Drugs with a High Risk of Generating Pharmacokinetic Interactions

DRUG	MECHANISM	EXAMPLES
Antacids	Reduced absorption	Antacids/tetracyclines
Bile acid sequestrants		Cholestyramine/digoxin
Proton pump inhibitors	Altered gastric pH	Ketoconazole absorption decreased
H ₂ -receptor blockers		
Rifampin	Induction of CYPs and/or P-glycoprotein	Decreased concentration and effects of
Carbamazepine		warfarin
Barbiturates		quinidine
Phenytoin		cyclosporine
St. John's wort		losartan
Glutethimide		oral contraceptives
Nevirapine (CYP3A; CYP2B6)		methadone, dabigatran
Tricyclic antidepressants	Inhibitors of CYP2D6	Increased effect of many β blockers
Fluoxetine		Decreased codeine effect; possible decreased tamoxifen effect
Quinidine		

DRUG	MECHANISM	EXAMPLES
Cimetidine	Inhibitor of multiple CYPs	Increased concentration and effects of warfarin theophylline phenytoin
Ketoconazole, itraconazole	Inhibitor of CYP3A	Increased concentration and toxicity of some HMG-CoA reductase inhibitors, colchicine
Erythromycin, clarithromycin		Cyclosporine, cisapride, terfenadine (now withdrawn)
Calcium channel blockers		Increased concentration and effects of indinavir (with ritonavir)
Ritonavir		Decreased clearance and dose requirement for cyclosporine (with calcium channel blockers)
Allopurinol	Xanthine oxidase inhibitor	Azathioprine and 6-mercaptopurine toxicity
Amiodarone	Inhibitor of many CYPs and of P-glycoprotein	Decreased clearance (risk of toxicity) for warfarin digoxin quinidine
Gemfibrozil (and other fibrates)	CYP3A inhibition	Rhabdomyolysis when co-prescribed with some HMG-CoA reductase inhibitors
Quinidine	P-glycoprotein inhibition	Risk of toxicity with P-glycoprotein substrates (e.g., digoxin, dabigatran)
Amiodarone		
Verapamil		
Cyclosporine		
Itraconazole		
Erythromycin		

DRUG	MECHANISM	EXAMPLES
Phenylbutazone	Inhibition of renal tubular transport	Increased risk of methotrexate toxicity with salicylates
Probenecid		
Salicylates		

PHARMACOKINETIC INTERACTIONS CAUSING DECREASED DRUG EFFECTS

Gastrointestinal absorption can be reduced if a drug interaction results in drug binding in the gut, as with aluminum-containing antacids, kaolin-pectin suspensions, or bile acid sequestrants. Drugs such as histamine H₂-receptor antagonists or proton pump inhibitors that alter gastric pH may decrease the solubility and hence absorption of weak bases such as [ketoconazole](#).

Expression of some genes responsible for drug elimination, notably *CYP3A* and *ABCB1*, can be markedly increased by inducing drugs, such as rifampin, carbamazepine, phenytoin, St. John's wort, and glutethimide, and by smoking, exposure to chlorinated insecticides, and chronic alcohol ingestion. Administration of inducing agents lowers plasma levels, and thus effects, over 2–3 weeks as gene expression is increased. If a drug dose is stabilized in the presence of an inducer that is subsequently stopped, major toxicity can occur as clearance returns to preinduction levels and drug concentrations rise. Individuals vary in the extent to which drug metabolism can be induced, likely through genetic mechanisms.

Interactions that inhibit the bioactivation of prodrugs will decrease drug effects ([Table 63-1](#)).

Interactions that decrease drug delivery to intracellular sites of action can decrease drug effects: tricyclic antidepressants can blunt the antihypertensive effect of clonidine by decreasing its uptake into adrenergic neurons. Reduced CNS penetration of multiple human immunodeficiency virus (HIV) protease inhibitors (with the attendant risk of facilitating viral replication in a sanctuary site) appears attributable to P-glycoprotein-mediated exclusion of the drug from the CNS; indeed, inhibition of P-glycoprotein has been proposed as a therapeutic approach to enhance drug entry to the CNS ([Fig. 63-5](#)).

PHARMACOKINETIC INTERACTIONS CAUSING INCREASED DRUG EFFECTS

The most common mechanism here is inhibition of drug elimination. In contrast to induction, new protein synthesis is not involved, and the effect develops as drug and any metabolites accumulate (a function of their elimination half-lives). Since shared substrates of a single enzyme can compete for access to the active site of the protein, many CYP substrates are also inhibitors. However, some drugs are especially potent as inhibitors (and occasionally may not even be substrates) of specific drug elimination pathways, and so it is in the use of these agents that clinicians must be most alert to the potential for interactions ([Table 63-2](#)).

Commonly implicated interacting drugs of this type include amiodarone, cimetidine, [erythromycin](#) and some other macrolide antibiotics (clarithromycin but not [azithromycin](#)), [ketoconazole](#) and other azole antifungals, the antiretroviral agent ritonavir, and high concentrations of grapefruit juice. The consequences of such interactions will depend on the drug whose elimination is being inhibited (see “The Concept of High-Risk Pharmacokinetics,” above). Examples include CYP3A inhibitors increasing the risk of [cyclosporine](#) toxicity or of rhabdomyolysis with some 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors ([lovastatin](#), [simvastatin](#), [atorvastatin](#), but not [pravastatin](#)), and P-glycoprotein inhibitors increasing the risk of toxicity with digoxin therapy or of bleeding with the [thrombin](#) inhibitor dabigatran.

These interactions can occasionally be exploited to therapeutic benefit. The antiviral ritonavir is a very potent CYP3A4 inhibitor that has been added to anti-HIV regimens, not because of its antiviral effects but because it decreases clearance, and hence increases efficacy, of other anti-HIV agents. Similarly, calcium channel blockers have been deliberately coadministered with [cyclosporine](#) to reduce its clearance and thus its maintenance dosage and cost.

Phenytoin, an inducer of many systems, including CYP3A, inhibits CYP2C9 and thus can reduce the bioactivation of losartan, with potential loss of antihypertensive effect, or the elimination of S-warfarin, with attendant increased bleeding risk.

Grapefruit (but not orange) juice inhibits CYP3A, especially at high doses; patients receiving drugs where even modest CYP3A inhibition may increase the risk of adverse effects (e.g., [cyclosporine](#), some HMG-CoA reductase inhibitors) should therefore avoid grapefruit juice.

CYP2D6 is markedly inhibited by quinidine, a number of neuroleptic drugs (chlorpromazine and haloperidol), and the selective serotonin reuptake inhibitors (SSRIs) fluoxetine and paroxetine. The clinical consequences of fluoxetine’s interaction with CYP2D6 substrates may not be apparent for weeks after the drug is started, because of its very long half-life and slow generation of a CYP2D6-inhibiting metabolite.

Azathioprine is metabolized to 6-mercaptopurine, which is then metabolized by thiopurine methyltransferase and by xanthine oxidase. When [allopurinol](#), an inhibitor of xanthine oxidase, is administered with standard doses of azathioprine or 6-mercaptopurine, life-threatening toxicity (bone marrow suppression) can result.

A number of drugs are secreted by the renal tubular transport systems for organic anions. Inhibition of these systems can cause excessive drug accumulation. Salicylate, for example, reduces the renal clearance of methotrexate, an interaction that may lead to methotrexate toxicity. Renal tubular secretion contributes substantially to the elimination of penicillin, which can be inhibited (to increase its therapeutic effect) by [probenecid](#). Similarly, inhibition of tubular cation transport by cimetidine decreases the renal clearance of dofetilide.

DRUG INTERACTIONS NOT MEDIATED BY CHANGES IN DRUG DISPOSITION

Drugs may act on separate components of a common process to generate effects greater than either has alone. While antithrombotic therapy with combinations of antiplatelet agents (glycoprotein IIb/IIIa inhibitors, aspirin, clopidogrel) and anticoagulants (e.g., warfarin, heparins, dabigatran, apixaban, rivaraxaban, edoxaban) is often used in the treatment of vascular disease, such combinations do carry an increased risk of bleeding.

Nonsteroidal anti-inflammatory drugs (NSAIDs) cause gastric ulcers, and in patients treated with oral anticoagulants, the risk of upper gastrointestinal bleeding is increased almost threefold by concomitant use of an NSAID.

[Indomethacin](#), piroxicam, and probably other NSAIDs antagonize the antihypertensive effects of β -adrenergic receptor blockers, diuretics, ACE inhibitors, and other drugs. The resulting elevation in blood pressure ranges from trivial to severe. This effect is not seen with aspirin and sulindac but has been found with the cyclooxygenase 2 (COX-2) inhibitor celecoxib.

Torsades de pointes ventricular tachycardia during administration of QT-prolonging antiarrhythmics (quinidine, sotalol, dofetilide) occurs much more frequently in patients receiving diuretics, probably reflecting hypokalemia. Low potassium not only prolongs the QT interval in the absence of drug but also potentiates drug block of ion channels that results in QT prolongation. Also, some diuretics have direct electrophysiologic actions that prolong QT.

The administration of supplemental potassium leads to more frequent and more severe hyperkalemia when potassium elimination is reduced by concurrent treatment with ACE inhibitors, spironolactone, eplerenone, amiloride, or triamterene.

The pharmacologic effects of sildenafil result from inhibition of the phosphodiesterase type 5 isoform that inactivates cyclic guanosine monophosphate (GMP) in the vasculature. [Nitroglycerin](#) and related nitrates used to treat angina produce vasodilation by elevating cyclic GMP. Thus, coadministration of these nitrates with sildenafil can cause profound hypotension, which can be catastrophic in patients with coronary disease.

Sometimes, combining drugs can increase overall efficacy and/or reduce drug-specific toxicity. Such therapeutically useful interactions are described in chapters dealing with specific disease entities.

ADVERSE DRUG REACTIONS

The beneficial effects of drugs are coupled with the inescapable risk of untoward effects. The morbidity and mortality from these adverse effects often present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. As well, some surveys have suggested that drug therapy for a range of chronic conditions such as psychiatric disease or hypertension does not achieve its desired goal in up to half of treated patients; thus, the most common “adverse” drug effect may be failure of efficacy.

Adverse reactions can be classified in two broad groups. Type A reactions result from exaggeration of an intended pharmacologic action of the drug, such as increased bleeding with anticoagulants or bone marrow suppression with some antineoplastics. Type B reactions result from toxic effects unrelated to the intended pharmacologic actions. The latter effects are often unanticipated (especially with new drugs) and frequently severe and may result from recognized (often immunologic) as well as previously undescribed mechanisms.

Drugs may increase the frequency of an event that is common in a general population, and this may be especially difficult to recognize; an excellent example is the increase in myocardial infarctions with the COX-2 inhibitor rofecoxib. Drugs can also cause rare and serious adverse effects, such as hematologic abnormalities, arrhythmias, severe skin reactions, or hepatic or renal dysfunction. Prior to regulatory approval and marketing (see below), new drugs are tested in relatively few patients who tend to be less sick and to have fewer concomitant diseases than those patients who subsequently receive the drug therapeutically. Because of the relatively small number of patients studied in clinical trials and the selected nature of these patients, rare adverse effects are generally not detected prior to a drug's approval; indeed, if they are detected, the new drugs are generally not approved. Therefore, physicians need to be cautious in the prescription of new drugs and alert for the appearance of previously unrecognized adverse events.

Elucidating mechanisms underlying adverse drug effects can assist development of safer compounds or allow a patient subset at especially high risk to be excluded from drug exposure. National adverse reaction reporting systems, such as those operated by the FDA (suspected adverse reactions can be reported online at <http://www.fda.gov/safety/medwatch/default.htm>) and the Committee on Safety of Medicines in Great Britain, can prove useful. The publication or reporting of a newly recognized adverse reaction can in a short time stimulate many similar such reports of reactions that previously had gone unrecognized.

Occasionally, “adverse” effects may be exploited to develop an entirely new indication for a drug. Unwanted hair growth during [minoxidil](#) treatment of severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in erectile tissue. These examples further reinforce the concept that prescribers must remain vigilant to the possibility that unusual symptoms may reflect unappreciated drug effects.

Some 25–50% of patients make errors in self-administration of prescribed medicines, and these errors can be responsible for adverse drug effects. Similarly, patients commit errors in taking OTC drugs by not reading or following the directions on the containers. Health care providers must recognize that providing directions with prescriptions does not always guarantee compliance.

In hospitals, drugs are administered in a controlled setting, and patient compliance is, in general, ensured. Errors may occur nevertheless—the wrong drug or dose may be given or the drug may be given to the wrong patient—and improved drug distribution and administration systems should help with this problem.

SCOPE OF THE PROBLEM

One estimate in the United Kingdom was that 6.5% of all hospital admissions are due to adverse drug reactions, and that 2.3% of these patients (0.15%) died as a result. The most common culprit drugs were aspirin, other NSAIDs, diuretics, warfarin, ACE inhibitors, antidepressants, opiates, digoxin, steroids, and clopidogrel. One study in the late 1990s suggested that adverse drug reactions were responsible for >100,000 in-hospital deaths in the United States, making them the 4th to 6th commonest cause of in-hospital death. Another study 10 years later showed no change in this trend.

In hospital, patients receive, on average, 10 different drugs during each hospitalization. The sicker the patient, the more drugs are given, and there is a corresponding increase in the likelihood of adverse drug reactions. When <6 different drugs are given to hospitalized patients, the probability of an adverse reaction is ~5%, but if >15 drugs are given, the probability is >40%. Serious adverse reactions are also well-recognized with “herbal” remedies and OTC compounds; examples include kava-associated hepatotoxicity, L-tryptophan-associated eosinophilia-myalgia, and phenylpropanolamine-associated stroke, each of which has caused fatalities.

TOXICITY UNRELATED TO A DRUG’S PRIMARY PHARMACOLOGIC ACTIVITY

Drugs or more commonly reactive metabolites generated by CYPs can covalently bind to tissue macromolecules (such as proteins or DNA) to cause tissue toxicity. Because of the reactive nature of these metabolites, covalent binding often occurs close to the site of production, typically the liver.

Acetaminophen

The most common cause of drug-induced hepatotoxicity is acetaminophen overdose ([Chap. 333](#)). Normally, reactive metabolites are detoxified by combining with hepatic glutathione. When glutathione becomes depleted, the metabolites bind instead to hepatic protein, with resultant hepatocyte damage. The hepatic necrosis produced by the ingestion of acetaminophen can be prevented or attenuated by the administration of substances such as *N*-acetylcysteine that reduce the binding of electrophilic metabolites to hepatic proteins. The risk of acetaminophen-related hepatic necrosis is increased in patients receiving drugs such as phenobarbital or phenytoin, which increase the rate of drug metabolism, or ethanol, which exhausts glutathione stores. Such toxicity has even occurred with therapeutic dosages, so patients at risk through these mechanisms should be warned.

Immunologic Reactions

Most pharmacologic agents are haptens, small molecules with low molecular weights (<2000) that are therefore poor immunogens. Generation of an immune response to a drug therefore often requires in vivo activation and covalent linkage to protein, carbohydrate, or nucleic acid.

Drug stimulation of antibody production may mediate tissue injury by several mechanisms. The antibody may attack the drug when the drug is covalently attached to a cell and thereby destroy the cell. This occurs in penicillin-induced hemolytic anemia. Antibody-drug-antigen complexes may be passively adsorbed by a

bystander cell, which is then destroyed by activation of complement; this occurs in quinine- and quinidine-induced thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by “paradoxical” thrombosis and is treated with **thrombin** inhibitors. Drugs or their reactive metabolites may alter a host tissue, rendering it antigenic and eliciting autoantibodies. For example, hydralazine and procainamide (or their reactive metabolites) can chemically alter nuclear material, stimulating the formation of antinuclear antibodies and occasionally causing lupus erythematosus. Drug-induced pure red cell aplasia (**Chap. 98**) is due to an immune-based drug reaction.

Serum sickness (**Chap. 345**) results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Complement activation occurs, chemotactic factors are generated locally, and an inflammatory response develops at the site of complex entrapment. Arthralgias, urticaria, lymphadenopathy, glomerulonephritis, or cerebritis may result. Foreign proteins (vaccines, streptokinase, therapeutic antibodies) and antibiotics are common causes. Many drugs, particularly antimicrobial agents, ACE inhibitors, and aspirin, can elicit anaphylaxis with production of IgE, which binds to mast cell membranes. Contact with a drug antigen initiates a series of biochemical events in the mast cell and results in the release of mediators that can produce the characteristic urticaria, wheezing, flushing, rhinorrhea, and (occasionally) hypotension.

Drugs may also elicit cell-mediated immune responses. One serious reaction is Steven-Johnson Syndrome/Toxic Epidermal Necrolysis (SJS/TEN), which can result in death due to T-cell-mediated massive skin sloughing. As described in **Chap. 64**, specific genetic variants appear necessary but not sufficient to elicit SJS/TEN. The mechanism is thought to be T cell activation by hapten-”self-peptide” interactions or direct binding of drug to HLA or T cell receptors.

DIAGNOSIS AND TREATMENT OF ADVERSE DRUG REACTIONS

The manifestations of drug-induced diseases frequently resemble those of other diseases, and a given set of manifestations may be produced by different and dissimilar drugs. Recognition of the role of a drug or drugs in an illness depends on appreciation of the possible adverse reactions to drugs in any disease, on identification of the temporal relationship between drug administration and development of the illness, and on familiarity with the common manifestations of the drugs.

A suspected adverse drug reaction developing after introduction of a new drug naturally implicates that drug; however, it is also important to remember that a drug interaction may be responsible. Thus, for example, a patient on a chronic stable warfarin dose may develop a bleeding complication after introduction of amiodarone; this does not reflect a direct reaction to amiodarone but rather its effect to inhibit warfarin metabolism. Many associations between particular drugs and specific reactions have been described, but there is always a “first time” for a novel association, and any drug should be suspected of causing an adverse effect if the clinical setting is appropriate.

Illness related to a drug's intended pharmacologic action is often more easily recognized than illness attributable to immune or other mechanisms.

For example, side effects such as cardiac arrhythmias in patients receiving digitalis, hypoglycemia in patients given insulin, or bleeding in patients receiving anticoagulants are more readily related to a specific drug than are symptoms such as rash, which may be caused by many drugs or by other factors. Drug fever often escapes initial diagnosis because fever is such a common manifestation of disease.

Electronic listings of adverse drug reactions can be useful. However, exhaustive compilations often provide little sense of perspective in terms of frequency and seriousness, which can vary considerably among patients.

Eliciting a drug history from each patient is important for diagnosis. Attention must be directed to OTC drugs and herbal preparations as well as to prescription drugs. Each type can be responsible for adverse drug effects, and adverse interactions may occur between OTC drugs and prescribed drugs. Loss of efficacy of oral contraceptives or [cyclosporine](#) with concurrent use of St. John's wort (a P-glycoprotein inducer) is an example. In addition, it is common for patients to be cared for by several physicians, and duplicative, additive, antagonistic, or synergistic drug combinations may therefore be administered if the physicians are not aware of the patients' drug histories. Every physician should determine what drugs a patient has been taking, for the previous month or two ideally, before prescribing any medications. Medications stopped for inefficacy or adverse effects should be documented to avoid pointless and potentially dangerous reexposure. A frequently overlooked source of additional drug exposure is topical therapy; for example, a patient complaining of bronchospasm may not mention that an ophthalmic beta blocker is being used unless specifically asked. A history of previous adverse drug effects in patients is common. Since these patients have shown a predisposition to drug-induced illnesses, such a history should dictate added caution in prescribing new drugs.

Laboratory studies may include demonstration of serum antibody in some persons with drug allergies involving cellular blood elements, as in agranulocytosis, hemolytic anemia, and thrombocytopenia. For example, both quinine and quinidine can produce platelet agglutination in vitro in the presence of complement and the serum from a patient who has developed thrombocytopenia following use of this drug. Biochemical abnormalities such as G6PD deficiency, serum pseudocholinesterase level, or genotyping may also be useful in diagnosis, especially after an adverse effect has occurred in the patient or a family member (see [Chap. 64](#)).

Once an adverse reaction is suspected, discontinuation of the suspected drug followed by disappearance of the reaction is presumptive evidence of a drug-induced illness. Confirming evidence may be sought by cautiously reintroducing the drug and seeing if the reaction reappears. However, that should be done only if confirmation would be useful in the future management of the patient and if the attempt would not entail undue risk. With concentration-dependent adverse reactions, lowering the dosage may cause the reaction to disappear, and raising it may cause the reaction to reappear. When the reaction is thought to be immunologic, however, readministration of the drug may be hazardous, since anaphylaxis may develop.

If the patient is receiving many drugs when an adverse reaction is suspected, the drugs likeliest to be responsible can usually be identified; this should include both potential culprit agents as well as drugs that alter their elimination. All drugs may be discontinued at once or, if this is not practical, discontinued one at a time, starting with the ones most suspect, and the patient observed for signs of improvement. The time needed for a concentration-dependent adverse effect to disappear depends on the time required for the concentration to fall below the range associated with the adverse effect; that, in turn, depends on the initial blood level and on the rate of elimination or metabolism of the drug. Adverse effects of drugs with long half-lives or those not directly related to serum concentration may take a considerable time to disappear.

THE DRUG DEVELOPMENT PROCESS

Drug therapy is an ancient feature of human culture. The first treatments were plant extracts discovered empirically to be effective for indications like fever, pain, or breathlessness. This symptom-based empiric approach to drug development was supplanted in the twentieth century by identification of compounds targeting more fundamental biologic processes, such as bacterial growth or elevated blood pressure. The term “magic bullet,” coined by Paul Ehrlich to describe the search for effective compounds for syphilis, captures the essence of the hope that understanding basic biologic processes will lead to highly effective new therapies.

A common starting point for the development of many widely used modern therapies has been basic biologic discovery that implicates potential target molecules: examples of such target molecules include HMG-CoA reductase, a key step in cholesterol biosynthesis, or the *BRAFV600E* mutation that appears to drive the development of some malignant melanomas and other tumors. The development of compounds targeting these molecules has not only revolutionized treatment for diseases such as hypercholesterolemia or malignant melanoma, but has also revealed new biologic features of disease. Thus, for example, initial spectacular successes with vemurafenib (which targets *BRAFV600E*) were followed by near-universal tumor relapse, strongly suggesting that inhibition of this pathway alone would be insufficient for tumor control. This reasoning, in turn, supports a view that many complex diseases will not lend themselves to cure by targeting a single magic bullet, but rather single drugs or combinations that attack multiple pathways whose perturbation results in disease. The use of combination therapy in settings such as hypertension, tuberculosis, HIV infection, and many cancers highlights the potential for such a “systems biology” view of drug therapy.

A common approach in contemporary drug development is to start with a high-throughput screening procedure to identify “lead” chemical(s) modulating the activity of a potential drug target. The next step is application of increasingly sophisticated medicinal chemistry-based modification of the “lead” to develop compounds with specificity for the chosen target, lack of “off-target” effects, and pharmacokinetic properties suitable for human use (e.g., consistent bioavailability, long elimination half-life, and no high-risk pharmacokinetic features). Drug evaluation in human subjects then proceeds from initial safety and tolerance (phase 1), dose finding (phase 2), and efficacy (phase 3). This is a very expensive process and the vast majority of lead compounds fail at some point. Thus, new approaches to identify likely successes and

failures early are needed. One idea, described further in [Chap. 64](#), is to use genomic and other high throughput profiling approaches not only to identify new drug targets but also to identify disease subsets for which drugs approved for other indications might be “repurposed” thereby avoiding the costly development process.

SUMMARY

Modern clinical pharmacology aims to replace empiricism in the use of drugs with therapy based on in-depth understanding of factors that determine an individual’s response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help guide further therapy with that drug or successors. This rapidly expanding understanding of variability in drug actions makes the process of prescribing drugs increasingly daunting for the practitioner. However, fundamental principles should guide this process:

The benefits of drug therapy, however defined, should always outweigh the risk.

The smallest dosage necessary to produce the desired effect should be used.

The number of medications and doses per day should be minimized.

Although the literature is rapidly expanding, accessing it is becoming easier; electronic tools to search databases of literature and unbiased opinion will become increasingly commonplace.

Genetics play a role in determining variability in drug response and may become a part of clinical practice.

Electronic medical record and pharmacy systems will increasingly incorporate prescribing advice, such as indicated medications not used; unindicated medications being prescribed; and potential dosing errors, drug interactions, or genetically determined drug responses.

Prescribers should be particularly wary when adding or stopping specific drugs that are especially liable to provoke interactions and adverse reactions.

Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READING

Landrigan CP et al: Temporal trends in rates of patient harm resulting from medical care. N Engl J Med 363:2124, 2010. [[PubMed: 21105794](#)]

Lazarou J et al: Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. JAMA 279:1200, 1998. [[PubMed: 9555760](#)]

MacRae CA et al: The future of cardiovascular therapeutics. Circulation 133:2610, 2016. [[PubMed: 27324356](#)]

Pirmohamed M et al: Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients. Br Med J 329:15, 2004.

Wheatley LM et al: Report from the National Institute of Allergy and Infectious Diseases workshop on drug allergy. J Allergy Clin Immunol 136:262, 2015. [[PubMed: 26254053](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 64: Pharmacogenomics

Dan M. Roden

FIGURE 64-1

INTRODUCTION

The previous chapter discussed mechanisms underlying variability in drug action, highlighting pharmacokinetic and pharmacodynamic pathways to beneficial and adverse drug events. Work in the past several decades has defined how genetic variation can play a prominent role in modulating these pathways. Initial studies described unusual drug responses due to single genetic variants in individual subjects, defining the field of pharmacogenetics. A more recent view extends this idea to multiple genetic variants across populations, and the term “pharmacogenomics” is often used. Understanding the role of genetic variation in drug response could improve the use of current drugs, avoid drug use in those at increased risk for adverse drug reactions (ADRs), guide development of new drugs, and even be used as a lens through which to understand mechanisms of diseases themselves. This chapter will outline the principles of pharmacogenomics, the evidence as currently available that genetic factors play a role in variable drug actions, and outline areas of controversy and future work.

PRINCIPLES OF GENETIC VARIATION AND DRUG RESPONSE

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members ([Chap. 457](#)) (See Also [Chaps. 456](#) and [457](#)). However, it is unusual for a drug response phenotype to be accurately measured in more than one family member, let alone across a kindred. Some clinical studies have examined drug disposition traits (such as urinary drug excretion after a fixed test dose) in twins, and have in some instances shown greater concordance in monozygotic compared to dizygotic pairs, supporting a genetic contribution to the trait under study. However, in general, non-family-based approaches are generally used to identify and validate DNA variants contributing to variable drug actions.

Types of Genetic Variants Influencing Drug Response

The commonest type of genetic variant is a single nucleotide polymorphism (SNP), and nonsynonymous SNPs (i.e., those that alter primary amino acid sequence encoded by a gene) are a common cause of variant function in genes regulating drug responses, often termed *pharmacogenes* ([Table 64-1](#)). Small insertions and deletions can similarly alter protein function, or lead to functionally important splice variation. Examples of synonymous coding region variants altering pharmacogene function have also been described; the postulated mechanism is an alteration in the rate of RNA translation, and hence in folding of the nascent protein. Variation in pharmacogene

promoters has been described, and copy number variation (gene deletion or multiple copies of the same gene) is also well described.

Table 64-1

Examples of Genetic Variation and Ancestry

Structural Variant	Example		Functional Effect	Minor Allele Frequency (%) ^a		
	Common Name	dbSNP		European	African	East Asian
Single nucleotide polymorphism (SNP) (or single nucleotide variant, SNV)	CYP2C9*2	rs1799853	R144C: Reduction of function	12.7	2.4	b
	CYP2C9*3	rs1057910	I359L: Loss of function	6.9	1.3	3.4
	CYP2C9*8	rs7900194	R150H: Reduction of function	b	5.6	b
	CYP2C19*2	rs4244285	Splicing defect: Loss of function	14.8	18.1	31.0
	CYP2C19*3	rs4986893	Premature stop: Loss of function	b	b	6.7
	CYP2C19*17	rs12248560	Gain of function	45	45	<5
	CYP2D6*4 ^c	rs3892097	Splicing defect: Loss of function	23.1	11.9	0.4
	CYP2D6*10 ^c	Multiple SNPs define CYP2D6*10 (reduction of function allele):				
		rs1065852	P34S	24.9	15.1	59.1
		rs1135840	S486T			
	CYP3A5*3	rs776746	Splicing defect: Loss of function	90	33	85
	VKORC1*2	rs9923231	Promoter variant associated with decreased warfarin dose	39	11	91

Structural Variant	Example		Functional Effect	Minor Allele Frequency (%) ^a		
	Common Name	dbSNP		European	African	East Asian
	VKORC1	rs61742245	D36Y: Reduction of function, associated with increased warfarin dose	5% in East Africa, Middle East, Oceania; rare elsewhere		
	ABCB1	rs1045642	Synonymous variant; may affect mRNA stability and protein folding	47.2	79.8	62.5
Insertion/deletion	UGT1A1*28		Reduction of function promoter variant (7 TA repeats versus 6 repeats in reference allele); homozygotes have Gilbert's syndrome	31.6	39.1	14.8
Multiple variants constituting specific haplotypes	HLA-B*15:01		Predispose to immunologically mediated adverse drug reactions	b	b	5
	HLA-B*57:01			6.8	1.0	1.6
Gene deletion	CYP2D6*5		Loss of function	2.7	6	5.6
Gene duplication	CYP2D6*1xN	Duplication of normal allele	Ultra-rapid metabolizer phenotype	0.8	1.5	0.3
				Up to 3% in North Africa and the Middle East		
	CYP2D6*4xN	Duplication of loss of function allele	Extensive or poor metabolizer phenotype, depending on the opposite allele	0.3	1.4	b

Note: Allele frequencies from <http://exac.broadinstitute.org/> and <https://cpicpgx.org/>.

^a

Includes heterozygotes and homozygotes.

b

Allele frequency <0.05%.

c

CYP2D6 is highly polymorphic and multiple SNPs may be required to define a specific variant. For example, rs1065852 is present in both *4 and *10 variants. See <http://www.cypalleles.ki.se>.

Table 64-1 lists examples of individual types of genomic variation and the impact they can have on function of pharmacogenes. Multiple genotyping approaches may be needed to detect important variants; for example, SNP assays may fail to detect large gene duplications, and highly polymorphic regions (such as human leukocyte antigens, HLA-B) are currently best evaluated by sequencing.

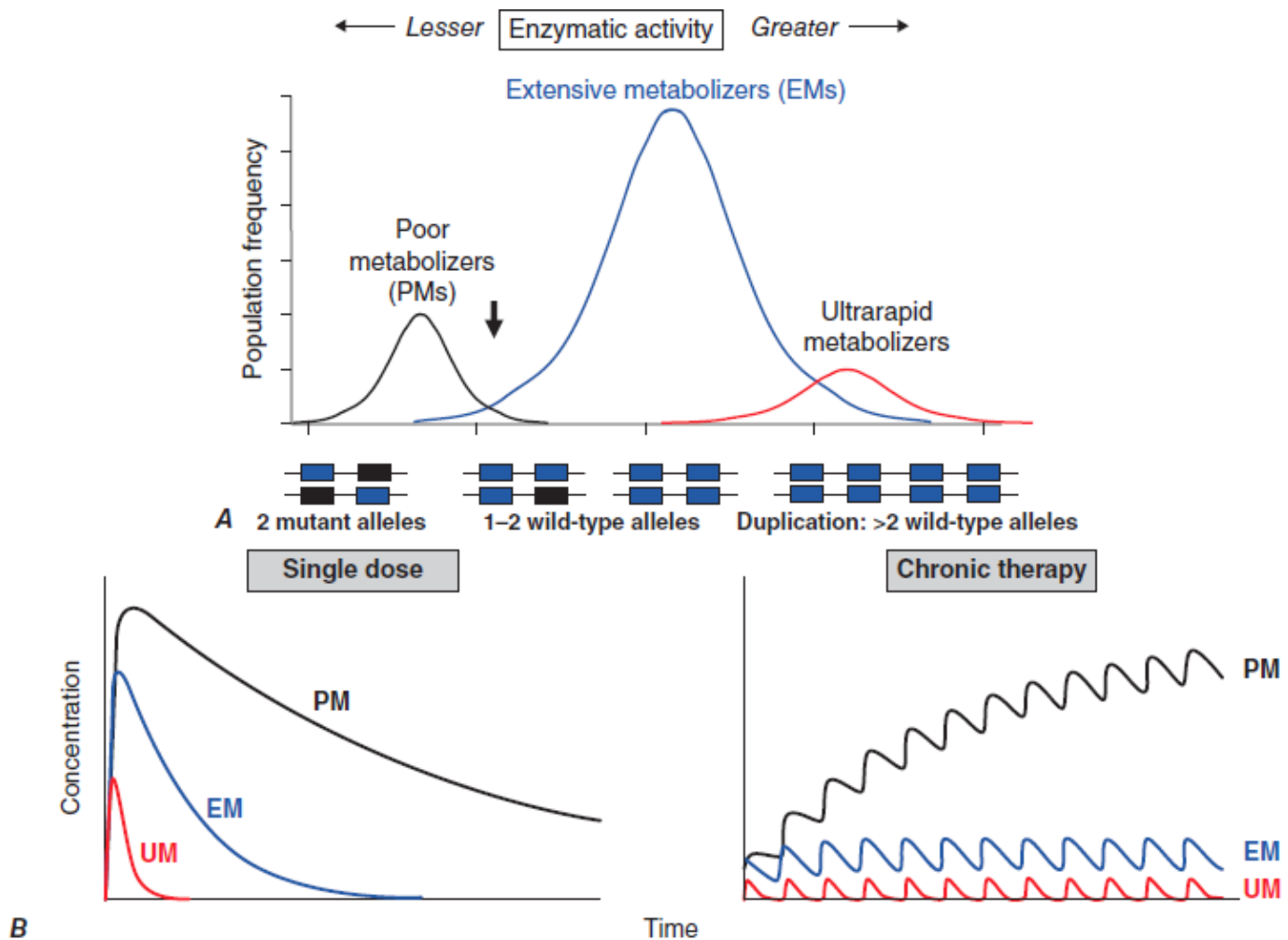
Table 64-1 highlights the fact that the frequency of important variation across pharmacogenes can vary strikingly by ancestry, with the result that certain ethnic groups may be at unusually high risk of displaying variant response to specific drugs.

Candidate Gene Approaches

Most studies to date have used an understanding of the molecular mechanisms modulating drug action to identify candidate genes in which variants could explain variable drug responses. One very common scenario is that variable drug actions can be attributed to variability in plasma drug concentrations. When plasma drug concentrations vary widely (e.g., more than an order of magnitude), especially if their distribution is non-unimodal as in **Fig. 64-1**, variants in single genes controlling drug concentrations often contribute. In this case, the most obvious candidate genes are those responsible for drug metabolism and elimination. Other candidate genes are those encoding the target molecules with which drugs interact to produce their effects or molecules modulating that response, including those involved in disease pathogenesis.

FIGURE 64-1

A. Distribution of CYP2D6 metabolic activity across a population. The heavy arrow indicates an antimode, separating poor metabolizer subjects (PMs, black), with two loss-of-function CYP2D6 alleles (black), indicated by the intron-exon structures below the chart. Individuals with one or two functional alleles are grouped together as extensive metabolizers (EMs, blue). Also shown are ultra-rapid metabolizers (UMs, red), with 2–12 functional copies of the gene, displaying the greatest enzyme activity. (*Adapted from M-L Dahl et al: J Pharmacol Exp Ther 274:516, 1995.*) **B.** These simulations show the predicted effects of CYP2D6 genotype on disposition of a substrate drug. With a single dose (left), there is an inverse “gene-dose” relationship between the number of active alleles and the areas under the time-concentration curves (smallest in UM subjects; highest in PM subjects); this indicates that clearance is greatest in UM subjects. In addition, elimination half-life is longest in PM subjects. The right panel shows that these single dose differences are exaggerated during chronic therapy: steady-state concentration is much higher in PM subjects (decreased clearance), as is the time required to achieve steady state (longer elimination half-life).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Genome-Wide Association Studies

The field has also had some success with “unbiased” approaches such as genome-wide association (GWA) ([Chap. 456](#)), particularly in identifying single variants associated with high risk for certain forms of drug toxicity. GWA studies have identified variants in the HLA-B locus that are associated with high risk for severe skin rashes during treatment with the anticonvulsant carbamazepine and hepatotoxicity with flucloxacillin, an antibiotic never marketed in the United States. A GWA study of simvastatin-associated myopathy identified a single noncoding SNP in *SLCO1B1*, encoding OATP1B1, a drug transporter known to modulate [simvastatin](#) uptake into the liver, which accounts for 60% of myopathy risk. GWA approaches have also implicated interferon variants in antileukemic responses and in response to therapy in hepatitis C. African-American subjects are known to have higher dose requirements to achieve stable anticoagulation with warfarin, due in part to variation in *CYP2C9* and *VKORC1*, discussed below. In addition, a GWA study identified novel SNPs near *CYP2C9* that contribute to this effect in African Americans.

GENETIC VARIANTS AFFECTING PHARMACOKINETICS

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition ([Table 64-2](#)). A distinct multimodal distribution of drug disposition (as shown in [Fig. 64-1](#)) argues for a predominant effect of variants in a single gene in the metabolism of that substrate. Individuals with two alleles (variants) encoding for nonfunctional protein make up one group, often termed *poor metabolizers* (PM phenotype). For most genes, many variants can produce such a loss of function, and assessing whether they are on the same or different alleles (i.e., the *diplotype*) can complicate the use of genotyping in clinical practice. Furthermore, some variants produce only partial loss of function, and the presence of more than one variant may be required to define a specific allele. Individuals with one functional allele, or multiple reduction of function alleles, make up a second (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (normal metabolizers, often termed *extensive metabolizers*, EMs). *Ultra-rapid metabolizers* (UMs) with especially high enzymatic activity (occasionally due to gene duplication; [Table 64-1](#) and [Fig. 64-1](#)) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways (see [Chap. 63](#), [Table 63-1](#)), and so EM individuals receiving such inhibitors can respond like PM patients (*phenocopying*). Polymorphisms in genes encoding drug uptake or drug efflux transporters may be other contributors to variability in drug delivery to target sites and, hence, in drug effects.

TABLE 64-2

Genetic Variants and Drug Responses

Gene	Drugs	Effect of Genetic Variants ^a
Variants in Drug Metabolism Pathways		
<i>CYP2C9</i>	losartan	Decreased bioactivation and effects (PMs)
	warfarin	Decreased dose requirements; possible increased bleeding risk (PMs)
	phenytoin	Decreased dose requirement (PMs)
<i>CYP2C19</i>	omeprazole, voriconazole	Decreased effect in EMs
	celecoxib	Exaggerated effect in PMs
	clopidogrel	Decreased effect in PMs and IMs Consider alternate drug in PMs and alternate drug or dose increase in IMs Possible increased bleeding risk in carriers of gain of function variants
	citalopram, escitalopram	Choose alternate drug in UMs; reduce dose in PMs
<i>CYP2D6</i>	codeine, tamoxifen	Decreased bioactivation and drug effects in PMs
	codeine	Respiratory depression in UMs
	tricyclic antidepressants ^b	Increased adverse effects in PMs: Consider dose decrease Decreased therapeutic effects in UMs: Consider alternate drug
	metoprolol, carvedilol, timolol, propafenone	Increased beta blockade in PMs
	Fluvoxamine	Reduce dose or chose alternate drug in PMs
<i>CYP3A5</i>	tacrolimus, vincristine	Decreased drug concentrations and effect (CYP3A5*3 carriers)

Gene	Drugs	Effect of Genetic Variants ^a
Dihydropyrimidine dehydrogenase (<i>DPYD</i>)	capecitabine, 5-fluorouracil, tegafur	Possible severe toxicity (PMs)
<i>NAT2</i>	rifampin, isoniazid , pyrazinamide, hydralazine, procainamide	Increased risk of toxicity in PMs
Thiopurine S-methyltransferase (<i>TPMT</i>)	azathioprine, 6-mercaptopurine, thioguanine	PMs: Increased risk of bone marrow aplasia EMs: Possible decreased drug action at usual dosages
Uridine diphosphate glucuronosyltransferase (<i>UGT1A1</i>)	irinotecan	PM homozygotes: Increased risk of severe adverse effects (diarrhea, bone marrow aplasia)
	atazanavir	High risk of hyperbilirubinemia during treatment; can result in drug discontinuation
Pseudocholinesterase (<i>BCHE</i>)	succinylcholine and other muscle relaxants	Prolonged paralysis (autosomal recessive). Diagnosis established by genotyping or by measuring serum cholinesterase activity.
Variants in Other Genes		
Glucose 6-phosphate dehydrogenase (<i>G6PD</i>)	rasburicase , primaquine, chloroquine	Increased risk of hemolytic anemia in G6PD-deficient subjects
HLA-B*15:02	carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN (mainly Asian subjects)
HLA-B*31:01	carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN and milder skin toxicities (Caucasian and Asian subjects)
HLA-B*15:02	phenytoin	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
HLA-B*57:01	abacavir	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
HLA-B*58:01	allopurinol	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
<i>IFNL3</i> (IL28B)	interferon	Variable response in hepatitis C therapy

Gene	Drugs	Effect of Genetic Variants ^a
<i>SLCO1B1</i>	simvastatin	Encodes a drug uptake transporter; variant non-synonymous single nucleotide polymorphism increases myopathy risk especially at higher dosages
<i>VKORC1</i>	warfarin	Decreased dose requirements with variant promoter haplotype Increased dose requirement in individuals with non-synonymous loss of function variants
<i>ITPA</i>	ribavirin	Variants modulate risk for hemolytic anemia
<i>RYR1</i>	general anesthetics	Variants predispose to malignant hyperthermia
<i>CFTR</i>	ivacaftor, lumacaftor	Targeted therapies for cystic fibrosis indicated only in certain genotypes
Variants in Other Genomes (Infectious Agents, Tumors)		
Chemokine C-C motif receptor (CCR5)	maraviroc	Drug effective only in HIV strains with CCR5 detectible
C-KIT	imatinib	In gastrointestinal stromal tumors, drug indicated only with c-kit-positive cases
ALK (anaplastic lymphoma kinase)	Crizotinib	Indicated in patients with non-small cell lung cancer and ALK mutations
Her2/neu overexpression	trastuzumab, lapatinib	Drugs indicated only with tumor overexpression
K-ras mutation	panitumumab, cetuximab	Lack of efficacy with <i>KRAS</i> mutation
Philadelphia chromosome	dasatinib, nilotinib, imatinib	Decreased efficacy in Philadelphia chromosome-negative chronic myelogenous leukemia

a

Drug effect in homozygotes unless otherwise specified.

b

Many tricyclic antidepressants and selective serotonin uptake inhibitors are metabolized by either CYP2D6, CYP2C19, or both, and some metabolites have pharmacologic activity. See <https://www.pharmgkb.org/view/dosing-guidelines.do>.

Note: EM, extensive metabolizer (normal enzymatic activity); IM, intermediate metabolizer (heterozygote for loss of function allele); PM, poor metabolizer (homozygote for reduced or loss of function allele); UM, ultra-rapid metabolizer (enzymatic activity much greater than normal, e.g., with gene duplication, [Fig. 64-1](#)). SJS/TEN: Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis.

Further data at:

U.S. Food and Drug Administration:

<http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm>

Pharmacogenetics Research Network/Knowledge Base: <http://www.pharmgkb.org>

The Clinical Pharmacogenomics Implementation Consortium: <https://www.pharmgkb.org/page/cpic>

CYP3A

Members of the CYP3A family (*CYP3A4*, *CYP3A5*) metabolize the greatest number of drugs in therapeutic use. CYP3A4 activity is highly variable (up to an order of magnitude) among individuals, but non-synonymous coding region polymorphisms (those that change the encoded amino acid) are rare. Thus, the underlying mechanism likely reflects genetic variation in regulatory regions.

Most subjects of European or Asian origin carry a polymorphism that disrupts splicing in the closely related *CYP3A5* gene. As a result, these individuals display reduced CYP3A5 activity whereas CYP3A5 activity tends to be greater in subjects of African origin. Decreased efficacy of the antirejection agent [tacrolimus](#) in subjects of African origin has been attributed to more rapid CYP3A5-mediated elimination and a lower risk of vincristine-associated neuropathy has been reported in CYP3A5 “expressers.”

CYP2D6

CYP2D6 is second to CYP3A4 in the number of commonly used drugs that it metabolizes. CYP2D6 activity is polymorphically distributed, with 5–10% of European- and African-derived populations (but very few Asians) displaying the PM phenotype ([Fig. 64-1](#)). Dozens of loss-of-function variants in *CYP2D6* have been described; the PM phenotype arises in individuals with two such alleles. In addition, ultra-rapid metabolizers with multiple functional copies of *CYP2D6* have been identified especially in East Africa, the Middle East, and Oceania. PMs have slower elimination rates and lower clearance of substrate drugs; as a consequence ([Fig. 64-1B](#)), steady state concentrations are higher and the time taken to achieve steady state is longer than in EMs ([see Chap. 63](#)). Conversely, UMs display very low steady state parent drug concentrations and an abbreviated time to steady state.

Codeine is biotransformed by CYP2D6 to the potent active metabolite morphine, so its effects are blunted in PMs and exaggerated in UMs. Deaths due to respiratory depression in children given codeine after tonsillectomy have been attributed to the UM trait, and the U.S. Food and Drug Administration (FDA) has revised the package insert to include a prominent “black box” warning against its use in this setting. In the case of drugs with beta-blocking

properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bronchospasm, bradycardia) are seen in PM subjects than in EMs. This can be seen not only with orally administered beta blockers such as [metoprolol](#) and carvedilol, but also with ophthalmic [timolol](#) and with the sodium channel–blocking antiarrhythmic propafenone, a CYP2D6 substrate with beta-blocking properties. Ultra-rapid metabolizers may require very high dosages of [nortriptyline](#) and other tricyclic antidepressants to achieve a therapeutic effect. Tamoxifen undergoes CYP2D6-mediated biotransformation to an active metabolite, so its efficacy may be in part related to this polymorphism. In addition, the widespread use of selective serotonin reuptake inhibitors (SSRIs) to treat tamoxifen-related hot flashes may also alter the drug's effects because many SSRIs, notably fluoxetine and paroxetine, are also CYP2D6 inhibitors.

CYP2C19

The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (2–3%) in other populations. The impact of polymorphic CYP2C19-mediated metabolism has been demonstrated with the proton pump inhibitor [omeprazole](#), where ulcer cure rates with “standard” dosages were much lower in EM patients (29%) than in PMs (100%). Thus, understanding the importance of this polymorphism would have been important in developing the drug, and knowing a patient's *CYP2C19* genotype should improve therapy. CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and several large retrospective studies have documented decreased efficacy (e.g., increased myocardial infarction after placement of coronary stents or increased stroke or transient ischemic attacks) among subjects with one or two reduction of function alleles. In addition, some studies suggest that [omeprazole](#) and possibly other proton pump inhibitors phenocopy this effect by inhibiting CYP2C19.

CYP2C9

There are common variants in *CYP2C9* that encode proteins with reduction or loss of catalytic function. These variant alleles are associated with increased rates of neurologic complications with phenytoin, hypoglycemia with glipizide, and reduced warfarin dose required to maintain stable anticoagulation. Rare patients homozygous for loss of function alleles may require very low warfarin dosages. Up to 50% of the variability in steady-state warfarin dose requirement is attributable to polymorphisms in *CYP2C9* and in the promoter of *VKORC1*, which encodes the warfarin target with lesser contributions by genes controlling vitamin K metabolism such as *CYP4F2*. The angiotensin-receptor blocker losartan is a prodrug that is bioactivated by CYP2C9; as a result, PMs and those receiving inhibitor drugs may display little response to therapy.

DPYD

Individuals homozygous for loss of function alleles in dihydropyrimidine dehydrogenase, encoded by DPYD, are at high risk for severe toxicity when exposed to the substrate anticancer drug 5-Fluorouracil (5-FU), as well as to capecitabine and tegafur, which are metabolized to 5-FU. Dose reductions have been recommended in intermediate metabolizers.

Transferase Variants

Thiopurine S-methyltransferase (TPMT) bioinactivates the antileukemic drug 6-mercaptopurine (6-MP) and 6-MP is itself an active metabolite of the immunosuppressive azathioprine. Homozygotes for alleles encoding inactive TPMT (1/300 individuals) predictably exhibit severe and potentially fatal pancytopenia on standard doses of azathioprine or 6-MP. On the other hand, homozygotes for fully functional alleles may display less anti-inflammatory or antileukemic effect with standard doses of the drugs.

N-acetylation is catalyzed by hepatic *N*-acetyl transferase (NAT), which represents the activity of two genes, *NAT1* and *NAT2*. Both enzymes transfer an acetyl group from acetyl coenzyme A to the drug; polymorphisms in *NAT2* are thought to underlie individual differences in the rate at which drugs are acetylated and thus define “rapid acetylators” and “slow acetylators.” Slow acetylators make up ~50% of European and African populations but are less common among East Asians. Slow acetylators have an increased incidence of the drug-induced lupus syndrome during procainamide and hydralazine therapy and of hepatitis with [isoniazid](#).

Individuals homozygous for a common promoter polymorphism that reduces transcription of uridine diphosphate glucuronosyltransferase (*UGT1A1*) have benign hyperbilirubinemia (Gilbert’s syndrome; [Chap. 330](#)). This variant has also been associated with diarrhea and increased bone marrow depression with the antineoplastic prodrug irinotecan, whose active metabolite is normally detoxified by UGT1A1-mediated glucuronidation. The antiretroviral atazanavir is a UGT1A1 inhibitor, and individuals with the Gilbert’s variant develop higher bilirubin levels during treatment. While this is benign, the hyperbilirubinemia can complicate clinical care because it may raise the question of whether coexistent hepatic injury is present.

Transporter Variants

The risk for myotoxicity with [simvastatin](#) and possibly other statins appears increased with variants in *SLCO1B1*. Variants in *ABCB1*, encoding the drug efflux transporter P-glycoprotein, may increase digoxin toxicity. Variants in the uptake transporters *MATE1* and *MATE2* have been reported to modulate [metformin](#)’s glucose-lowering activity.

GENETIC VARIANTS AFFECTING PHARMACODYNAMICS

A variant in the *VKORC1* promoter, especially common in Asian subjects ([Table 64-1](#)), reduces transcriptional activity and warfarin dose requirement. Multiple polymorphisms identified in the β_2 -adrenergic receptor appear to be linked to specific phenotypes in asthma and congestive heart failure, diseases in which β_2 -receptor function might be expected to determine prognosis. Polymorphisms in the β_2 -receptor gene have also been associated with response to inhaled β_2 -receptor agonists, while those in the β_1 -adrenergic receptor gene have been associated with variability in heart rate slowing and blood pressure lowering. In addition, in heart failure, the arginine allele of the common β_1 -adrenergic receptor gene polymorphism R389G has been associated with decreased mortality and decreased incidence of atrial fibrillation during treatment with the investigational beta blocker bucindolol.

Drugs may also interact with genetic pathways of disease to elicit or exacerbate symptoms of the underlying conditions. In the porphyrias, CYP inducers are thought to increase the activity of enzymes proximal to the

deficient enzyme, exacerbating or triggering attacks ([Chap. 409](#)). Deficiency of glucose-6-phosphate dehydrogenase (G6PD), most often in individuals of African, Mediterranean, or South Asian descent, increases the risk of hemolytic anemia in response to the antimalarial primaquine ([Chap. 96](#)) and the uric acid-lowering agent [rasburicase](#), which does not cause hemolysis in patients with normal amounts of the enzyme. Patients with mutations in *RYR1* encoding the skeletal muscle intracellular release calcium (also termed type 1 ryanodine receptor) are asymptomatic until exposed to certain general anesthetics, which can trigger the rare syndrome of malignant hyperthermia. Certain antiarrhythmics and other drugs can produce marked QT prolongation and torsades de pointes ([Chap. 241](#)), and in some patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

Immunologically Mediated Drug Reactions

The Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are potentially fatal skin reactions now increasingly recognized to be linked to specific HLA alleles (see [Table 64-2](#)). Some cases of hepatotoxicity have also been linked to variants in this region. The frequency of risk alleles often varies by ancestry ([Table 64-1](#)). The HLA risk alleles appear to be necessary but not sufficient to elicit these reactions. For example, HLA-B*57:01 is a risk allele for abacavir-related SJS/TEN and flucloxacillin-related hepatotoxicity. However, while 55% of abacavir-exposed subjects will develop reaction, only 1/10,000 subjects exposed to flucloxacillin develop hepatotoxicity. Thus, a third factor, the nature of which has not yet been established, seems necessary.

Tumor and Infectious Agent Genomes

The actions of drugs used to treat infectious or neoplastic disease may be modulated by variants in these nonhuman germline genomes. Genotyping tumors is a rapidly evolving approach to target therapies to underlying mechanisms and to avoid potentially toxic therapy in patients who would derive no benefit ([Chap. 67](#)). Trastuzumab, which potentiates anthracycline-related cardiotoxicity, is ineffective in breast cancers that do not express the herceptin receptor. Imatinib targets a specific tyrosine kinase, BCR-Abl1, that is generated by the translocation that creates the Philadelphia chromosome typical of chronic myelogenous leukemia (CML). BCR-Abl1 is not only active but may be central to the pathogenesis of CML; use of imatinib and other BCR-Abl1 inhibitors has resulted in remarkable efficacy not only in CML but also in other BCR-Abl1-positive tumors such as gastrointestinal stromal tumors (see [Chap. 67](#)). Similarly, the anti-epidermal growth factor receptor (EGFR) antibodies cetuximab and panitumumab appear especially effective in colon cancers in which K-ras, a G protein in the EGFR pathway, is not mutated. Vemurafenib does not inhibit wild-type *BRAF* but is active against the V600E mutant form of the kinase. Crizotinib is highly effective in non-small cell lung cancers harboring anaplastic lymphoma kinase (ALK) mutations.

INCORPORATING PHARMACOGENETIC INFORMATION INTO CLINICAL PRACTICE

The discovery of common variant alleles with relatively large effects on drug response raises the prospect that these variants could be used to guide therapy. Desired outcomes could be better ways of choosing likely effective drugs and dosages, or avoiding drugs that are likely to produce severe adverse drug events or be ineffective in individual subjects. Indeed, the FDA now incorporates pharmacogenetic data into package inserts meant to guide

prescribing. A decision to adopt pharmacogenetically guided dosing for a given drug depends on multiple factors. The most important are the magnitude and clinical importance of the genetic effect and the strength of evidence linking genetic variation to variable drug effects (e.g., anecdote versus post-hoc analysis of clinical trial data versus randomized clinical trial, RCT). The evidence can be strengthened if statistical arguments from clinical trial data are complemented by an understanding of underlying physiologic mechanisms. Cost versus expected benefit may also be a factor.

Reactive versus Preemptive Approaches

Two approaches to pharmacogenetic implementation have been put in place at both “early adopter” institutions and are currently being evaluated. In the first, variant-specific assays are ordered at the time of drug prescription and delivered rapidly (often within an hour or two) and the results then used to guide therapy with that specific drug. The alternative to this “reactive” approach is a “preemptive” approach in which pharmacogenetic testing for large numbers of potential variants across many drugs is undertaken prior to prescription of any such drug. The data are then available in electronic health record (EHR) systems and coupled to real time clinical decision support (CDS). When a drug whose effects are known to be influenced by pharmacogenetic variants is prescribed, the EHR system looks up whether variants likely to affect response are present; if so, CDS will alert healthcare providers that an alternate drug or a different dose may be required.

Challenges

There are multiple challenges in putting in place either system. Assay validity and reproducibility have been issues in the past, but are less likely now. National consortia are now being put in place to develop standards for pharmacogenetic CDS. While common variants in genes such as those listed in [Table 64-1](#) have been clearly associated with variable drug responses, the effect of rare variants, now readily discoverable by large scale sequencing, is unknown. The extent to which a dose adjustment might be recommended may vary depending on whether zero, one, or two variant alleles are present, and whether such variants are reduction of function, loss-of-function, or gain of function. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has developed and published guidelines for multiple drug-gene pairs focusing on the question of what might be an appropriate drug dose adjustment given the availability of genetic data. CPIC does not, however, address the question of when or how such genetic testing should be undertaken.

Developing Evidence that Pharmacogenetic Testing Alters Drug Outcomes

A major issue is whether pharmacogenetic testing affects important drug response outcomes. When the evidence is compelling, alternate therapies are not available, and there are clear recommendations for dosage adjustment in subjects with variants, there is a strong argument for deploying genetic testing as a guide to prescribing; HLA-B*57:01 testing for abacavir is an example described below. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the drug effect may be amenable to monitoring by other approaches.

One school argues that the physiology and pharmacology are known, and that RCTs are, therefore, unnecessary (and conceivably unethical). The analogy is sometimes drawn to well-recognized dose adjustment of renally excreted drugs in the presence of renal dysfunction. RCTs have not been conducted and the idea of such dose

adjustment is well accepted in the medical community and recommended in FDA-approved drug labels. Others have argued that the effect of genetic variants is generally modest and variability in drug actions has many non-genetic sources, so genetic testing might provide marginal benefit at best.

Efforts to demonstrate the value of pharmacogenetic testing have met with mixed results. An RCT clearly showed that HLA-B*57:01 testing eliminates SJS/TEN due to abacavir. Similarly, regulatory authorities in some countries in Southeast Asia mandated HLA-B*15:02 testing prior to initiation of carbamazepine; however, in this case, an unfortunate outcome was that while the use of carbamazepine dropped, it was often substituted by phenytoin (another drug associated with SJS/TEN), so the incidence of the severe ADR was unchanged.

RCTs evaluating the effect of using pharmacogenetically guided therapy to optimize warfarin treatment have shown either no effect or a modest benefit of incorporating genetic information into prescribing the drug. These RCTs focused on time in therapeutic range in the first 4–12 weeks of treatment, and were not powered to examine outcomes such as recurrent thrombosis or bleeding. Retrospective analyses of bleeding cases vs non-bleeding controls in EHRs and administrative databases have suggested a role for CYP2C9*3 or the variants in V433M variant in CYP4F2 in mediating this risk.

While large retrospective analyses indicate that CYP2C19 loss of function variants decrease clopidogrel efficacy, RCTs are difficult to design: many argue that it is unethical to randomize individuals known to be homozygous for loss of function alleles, since administering clopidogrel is then tantamount to administering placebo. However, trials examining outcomes in only heterozygotes might require very large numbers of subjects.

New effective alternate therapies to warfarin and clopidogrel that appear to lack important pharmacogenetic variants have emerged. One approach to therapy, therefore, is to use pharmacogenetic testing to identify subjects in whom variants are absent and therefore standard doses of the conventional inexpensive drugs are likely to be effective and reserve alternate more expensive therapies for subjects likely to have variant responses to warfarin or clopidogrel. As price drops and as experience grows with newer agents, it is likely that clopidogrel and warfarin will be largely supplanted.

GENETICS AND DRUG DEVELOPMENT

Genetic tools are now being increasingly used to identify or validate new drug targets. Initial studies in this field suggest that a new drug development program is more likely to succeed if evidence from human genetics supports the role of a possible drug target in disease pathogenesis and suggests that the risk of toxicity due to high risk pharmacokinetics or other mechanisms is small.

Finding Protective Alleles Can Identify Drug Targets

One example of using genetics to identify a new drug target started with the discovery that very rare gain of function variants in *PCSK9* are a rare cause of familial hypercholesterolemia. Subsequently, population studies showed that carriers of loss of function SNPs (2.5% of African Americans) had decreased low-density lipoprotein, decreased incidence of coronary artery disease, and no deleterious consequences in other organ systems. These data triggered the development of PCSK9 antagonists which were marketed less than 10 years after the initial population studies. Other targets implicated by similar population genetic studies include SLC30A8 for the

prevention of type 2 diabetes and APOC3 for hypertriglyceridemia. In the latter examples, the identification of an apparently protective effect of rare loss-of-function alleles required very large datasets (>100,000) coupling DNA to longitudinal clinical information; long-term epidemiologic studies like the Framingham Heart Study or EHR systems are now being harnessed to address this opportunity.

Cancer

In cancer, tumor sequencing has identified new targets for drug development, often constitutively active kinases. A problem in this area has been the rapid emergence of drug resistance, often after extraordinary initial responses. For example, 40% of melanomas appear to be driven by the V600E mutant form of BRAF, and the specific inhibitor vemurafenib can produce clinically spectacular remission. However, durable responses are rare, and it is now apparent that combination therapy, often with inhibitors of the MEK pathway, can provide improved therapy. Another approach that is rapidly gaining wide use in cancer are drugs that reverse immune system inhibition (**Chap. 69**). In some patients the release of this “break” can provide durable remissions, whereas in others, severe adverse events, including colitis, pneumonitis, and myocarditis, have been reported. Understanding the mechanisms underlying variability to these therapies is a major emerging challenge in the field.

Using Multiple Data Types

The development of methods to understand associations across multiple large datasets is another approach that is being explored in drug development. For example, a GWA of risk of rheumatoid arthritis identified multiple risk loci and many encode proteins that are known targets for intervention in the disease. Interestingly, others encode proteins that are targets for drugs used in other conditions, such as certain cancers, raising the question of whether such drugs could be “repurposed” for rheumatoid arthritis. An extension of this approach is the broader issue of systems pharmacology, in which multiple sources of data are used to identify potential molecules or pathways that would be amenable to treatment, by new drugs or by existing agents, using analysis of genomic, transcriptomic, proteomic, and other large datasets. Similar approaches are being developed to predict toxicity expected from targeting specific genes or disease pathways.

SUMMARY

The science of pharmacogenomics has evolved from isolated examples of rare adverse drug actions to a more comprehensive view of the role of genetic variation in mediating the effects of most drugs. Current principles include:

Genetic variants with an important effect on drug actions can be common and their frequencies often vary by ancestry.

One common mechanism is modulation of drug concentrations.

No practitioner can be expected to remember all variants important for all drugs. Electronic data systems can now be accessed to describe this information. Ultimately, this information will be used by linking individual pharmacogenetic data to smart electronic health record systems.

Incorporating genetic approaches into drug development projects hold the promise of more rapid development of targeted, safe, and effective therapies.

FURTHER READING

Mallal S et al: HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med 358:568, 2008. [[PubMed: 18256392](#)]

Nelson MR et al: The support of human genetic evidence for approved drug indications. Nat Genet 47:856, 2015. [[PubMed: 26121088](#)]

Relling MV, Evans WE: Pharmacogenomics in the clinic. Nature 526:343, 2015. [[PubMed: 26469045](#)]

Wang L et al: Genomics and drug response. N Engl J Med 364:1144, 2011. [[PubMed: 21428770](#)]

Weeke P, Roden DM: Applied pharmacogenomics in cardiovascular medicine. Annu Rev Med 65:81, 2014. [[PubMed: 24111889](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 116: Molecular Mechanisms of Microbial Pathogenesis

Gerald B. Pier

INTRODUCTION

Over the past five decades, molecular studies of the pathogenesis of microorganisms have yielded a torrent of information about the various microbial and host molecules that contribute to the processes of infection and disease. These processes can be classified into several stages: microbial encounter with and entry into the host; microbial growth after entry; avoidance of innate host defenses; tissue invasion and tropism; tissue damage; and transmission to new hosts. *Virulence* is the measure of an organism's capacity to cause disease and is a function of the pathogenic factors elaborated by microbes. These factors promote *colonization* (the simple presence of potentially pathogenic microbes in or on a host), *infection* (attachment and growth of pathogens and avoidance of host defenses), and *disease* (often, but not always, the result of activities of secreted toxins or toxic metabolites). In addition, the host's inflammatory response to infection greatly contributes to disease and its attendant clinical signs and symptoms. A recent explosion of interest in the *microbiome* (the collection of microbial genomes present in or on mammalian organisms) and the *microbiota* (the collection of microbes residing in and on mammalian organisms) and their impact on physiology of, susceptibility to, and response to infection and on immune system development has greatly expanded our understanding of host–pathogen interactions. Furthermore, investigations in this field have documented effects of the microbiome on all aspects of animal—and even plant—physiology, greatly increasing our knowledge of the everyday influence of host–microbe interactions on life.

MICROBIAL ENTRY AND ADHERENCE

The Microbiome

We now know that the indigenous microbial organisms living in close association with almost all animals and plants are organized into complex communities that strongly modulate overall host physiology, including the ability of pathogenic microbes to establish themselves in or on host surfaces. The sheer numbers of these microbes and their genomic variability often exceed the numbers of host cells and the variability of host genes in a typical animal. Changes and differences in microbiomes within and between individuals, currently characterized by high-throughput DNA sequencing techniques and bioinformatic analysis, impact such diverse conditions as obesity; type 1 diabetes; cognition; neurologic states; autoimmune diseases; skin, gastrointestinal, respiratory, and vaginal infectious diseases; and development and control of the immune system. It has been difficult to directly associate specific types of microbiomes with pathophysiologic states, and our understanding of the degree to which microbial species are conserved or variable within human and other animal microbiotas is evolving. Experimental studies in laboratory animals, particularly in germ-free mammals, show the potent ability of changes in the microbiota to manipulate health status and outcomes. One of the clearest functions of the microbiota is to influence and mature the cells of the immune system, thereby exerting a major effect on susceptibility and resistance to microbial infection. The degree to which studies of the microbiome will translate into strategies for the management of human health and disease (e.g., the use of fecal transplants to treat and prevent recurrences of serious *Clostridium difficile* infection) is still an open question. For the moment, defining clusters of organisms associated with diseases may be more feasible than identifying single organisms or microbial molecules. Results from the Human Microbiome Project suggest a high level of variability among individuals in microbiome components, although many individuals appear to maintain a fairly conserved microbiome throughout their lives. In the context of infectious diseases, changes and disruptions of the indigenous microbiome—i.e., alterations of the normal flora due to antibiotic and immunosuppressive drug use, environmental changes, and the effects of microbial virulence factors used to displace the indigenous microbial flora and thus to facilitate pathogen colonization—have a strong and often fundamental impact on the progression of infection. While the technology for defining and understanding the microbiome is still quite young, there is little doubt that the resulting data will markedly affect our concepts of and approaches to microbial pathogenesis and infectious disease treatment.

Entry Sites

A microbial pathogen can potentially enter any part of a host organism. In general, the type of disease produced by a particular microbe is often a direct consequence of its route of entry into the body. The most common sites of entry are mucosal surfaces (the respiratory, alimentary, and urogenital tracts) and the skin. Ingestion, inhalation, and sexual contact are typical routes of microbial entry. Other portals of entry include sites of skin injury (cuts, bites, burns, trauma) along with injection via natural (e.g., vector-borne) or artificial (e.g., needlestick injury) routes. A few pathogens, such as *Schistosoma* species, can penetrate unbroken skin. The conjunctiva can serve as an entry point for pathogens of the eye, which occasionally spread systemically from that site.

Microbial entry usually relies on the presence of specific factors needed for persistence and growth in a tissue. Fecal–oral spread via the alimentary tract requires a biologic profile consistent with survival in the varied environments of the gastrointestinal tract (including the low pH of the stomach and the high bile content of the intestine) as well as in contaminated food or water outside the host. Organisms that gain entry via the respiratory tract survive well in small moist droplets produced during sneezing and coughing. Pathogens that enter by venereal routes often survive best in the warm moist environment of the urogenital mucosa and have restricted host ranges (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, and HIV).

The biology of microbes entering through the skin is highly varied. Some of these organisms can survive under a broad range of environmental conditions, such as those in the salivary glands or alimentary tracts of arthropod vectors, the mouths of larger animals, soil, and water. A complex biology allows protozoan parasites such as *Plasmodium*, *Leishmania*, and *Trypanosoma* species to undergo morphogenic changes that permit transmission of the organism to mammalian hosts during insect feeding for blood meals. Plasmodia are injected as infective sporozoites from the salivary glands during mosquito feeding. *Leishmania* parasites are regurgitated as promastigotes from the alimentary tract of sandflies and injected by bite into a susceptible host. Trypanosomes are first ingested from infected hosts by reduviid bugs; the pathogens then multiply in the gastrointestinal tract of the insects and are released in feces onto the host's skin during subsequent feedings. Most microbes that land directly on intact skin are destined to die, as survival on the skin or in hair follicles requires resistance to fatty acids, low pH, and other antimicrobial factors on the skin. Once it is damaged (and particularly if it becomes necrotic), the skin can be a major portal of entry and growth for pathogens and elaboration of their toxic products. Burn wound infections and tetanus are clear examples. After animal bites, pathogens resident in the animal's saliva gain access to the victim's tissues through the damaged skin. Rabies is the paradigm for this pathogenic process; rabies virus grows in striated muscle cells at the site of inoculation.

Microbial Adherence

Once in or on a host, many microbes must situate themselves favorably to avoid clearance mechanisms, in part by microbial anchoring to a tissue or tissue factor. (One possible exception is an organism that directly enters the bloodstream and multiplies there.) Because most host cells—responding to activation of innate immunity (see “[Avoidance of Innate Host Defenses](#),” below)—express multiple surface and cytoplasmic molecules that detect pathogens and pathogen factors, a complex interplay ensues and determines whether the microbe will avoid host clearance and remain in a tissue. Viruses and intracellular pathogens like *Mycobacterium tuberculosis* must bind to cells and enter them, whereas common extracellular bacterial pathogens of the human respiratory tract survive better if they avoid binding to pulmonary epithelial cells.

Specific ligands or adhesins for host receptors constitute a major area of study in microbial pathogenesis. Adhesins comprise a wide range of surface structures, anchoring the microbe to a tissue and promoting cellular entry as well as eliciting host responses critical to innate immunity ([Table 116-1](#)). Most microbes produce multiple adhesins specific for multiple host receptors that often are redundant, are serologically variable, and act additively or synergistically with other microbial factors to promote sticking to host tissues. In addition, some microbes adsorb host proteins onto their surface and use the natural host protein receptor for binding and entry into cells. While it is clear that, for some pathogenic organisms, blocking adherence can be a means to prevent infection, for others it could have unintended consequences, decreasing the innate host response that facilitates elimination of the infecting microbe.

Examples of Microbial Ligand–Receptor Interactions

MICROORGANISM	TYPE OF MICROBIAL LIGAND	HOST RECEPTOR
Viral Pathogens		
Influenza virus	Hemagglutinin	Sialic acid
Measles virus		
Vaccine strain	Hemagglutinin	CD46/moesin/signaling lymphocytic activation molecule (SLAM)/nectin-4
Wild-type strains	Hemagglutinin	
Human herpesvirus type 6A	Glycoprotein complex gH/gL/gQ1/gQ2	CD46
Herpes simplex virus	Glycoprotein C	Heparan sulfate
HIV	Surface glycoprotein	CD4 and chemokine receptors (CCR5 and CXCR4)
Epstein-Barr virus	Envelope protein	CD21 (CR2)
Adenovirus	Fiber protein	Coxsackie-adenovirus receptor (CAR)
Coxsackievirus	Viral coat proteins	CAR and major histocompatibility class I antigens
Bacterial Pathogens		
<i>Neisseria</i> spp.	Pili	Membrane cofactor protein (CD46)
<i>Pseudomonas aeruginosa</i>	Pili and flagella Lipopolysaccharide	Asialo-GM1 Cystic fibrosis transmembrane conductance regulator (CFTR)
<i>Escherichia coli</i>	Pili	Ceramides/mannose and digalactosyl residues
<i>Streptococcus pyogenes</i>	Hyaluronic acid capsule	CD44
<i>Yersinia</i> spp.	Invasin/accessory invasin locus	β_1 Integrins
<i>Bordetella pertussis</i>	Filamentous hemagglutinin	CR3
<i>Legionella pneumophila</i>	Adsorbed C3bi	CR3
<i>Mycobacterium tuberculosis</i>	Adsorbed C3bi	CR3; DC-SIGN
Fungal Pathogens		
<i>Blastomyces dermatitidis</i>	WI-1	Possibly matrix proteins and integrins
<i>Candida albicans</i>	Int1p	Extracellular matrix proteins
Protozoal Pathogens		
<i>Plasmodium vivax</i>	Merozoite form	Duffy Fy antigen
<i>Plasmodium falciparum</i>	Erythrocyte-binding protein 175 (EBA-175)	Glycophorin A
<i>Entamoeba histolytica</i>	Surface lectin	N-Acetylglucosamine

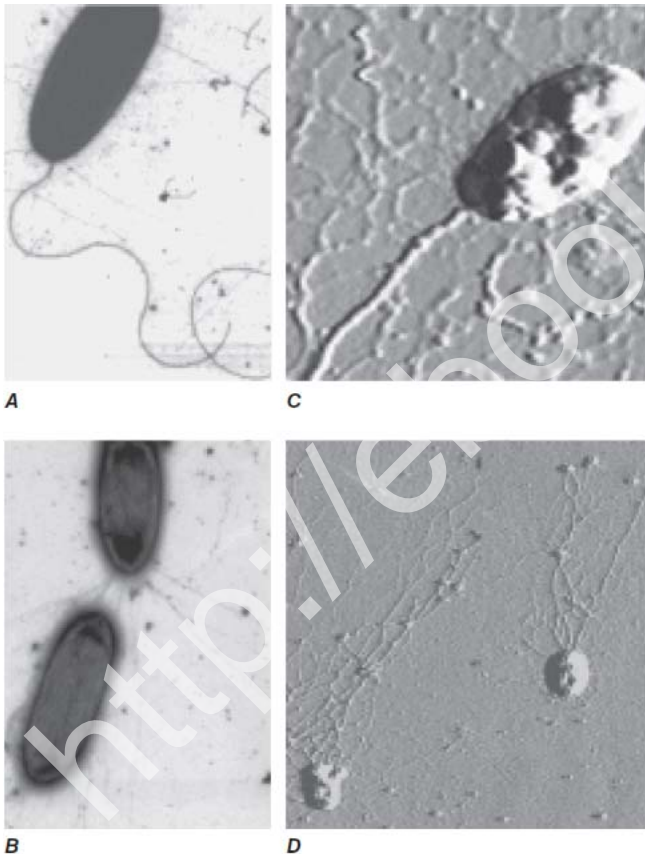
All viral pathogens must bind to host cells, enter them, and replicate within them. Viral coat proteins serve as ligands for cellular entry, and more than one ligand–receptor interaction may be needed. In some types of viruses, such as lipid bilayer–encapsulated Retroviridae or Rhabdoviridae, a single protein mediates both viral binding and entry via fusion with the host cell membrane. In other cases, a second viral fusion protein is needed to complete viral entry. HIV uses its envelope glycoprotein (gp) 120 to enter host cells by binding to both CD4 and one of two receptors for chemokines (CCR5 or CXCR4). Measles virus requires two proteins for cellular entry: the hemagglutinin (H) glycoprotein of wild-type measles virus binds to the signaling lymphocytic activation molecule (SLAM or CD150) on macrophages and dendritic cells, where the virus initially replicates, and also to nectin-4 on respiratory epithelial cells, where later replication occurs. The vaccine strain of measles virus binds to both CD46 and SLAM. For full cellular entry, however, measles virus requires a second fusion (F) protein. The gB and gC proteins on herpes simplex virus bind to heparan sulfate, although this adherence is not essential for entry but rather serves to concentrate virions close to the cell surface; this step is followed by attachment to mammalian cells mediated by the viral gD protein, with subsequent formation of a homotrimer of viral gB protein or a heterodimer of viral gH and gL proteins that permits fusion of the viral envelope with the host cell membrane. Herpes simplex virus can use a number of eukaryotic cell-surface receptors for entry, including the herpesvirus entry mediator, members of the immunoglobulin superfamily, the proteins nectin-1 and nectin-2, and modified heparan sulfate.

Bacterial Adhesins

Among the adhesins studied in greatest detail are bacterial pili and flagella (Fig. 116-1). *Pili* or *fimbriae* are commonly used by gram-negative bacteria for attachment to host cells and tissues; similar factors are produced by gram-positive organisms such as group B streptococci. In electron micrographs, these hairlike projections (up to several hundred per cell) may be confined to one end of the organism (*polar pili*) or distributed more evenly over the surface. An individual cell may have pili with a variety of functions. Most pili are made up of a major pilin protein subunit (17,000–30,000 Da) that polymerizes to form the pilus. Many strains of *Escherichia coli* isolated from urinary tract infections express a mannose-binding type 1 pilus that attaches to the uroplakins coating the cells in the bladder epithelium. Other strains produce the Pap (pyelonephritis-associated) or P pilus adhesin that mediates binding to digalactose (gal-gal) residues on globosides of the human P blood groups. Both of these pili have proteins located at the tips of the main pilus unit that are critical to the binding specificity of the whole pilus unit. *E. coli* cells causing diarrheal disease express pilus-like receptors for enterocytes on the small bowel, along with other receptors termed *colonization factors*.

FIGURE 116-1

Bacterial surface structures. A and B. Traditional electron micrographic images of fixed cells of *Pseudomonas aeruginosa*. Flagella (**A**) and pili (**B**) project out from the bacterial poles. **C and D.** Atomic force microscopic image of live *P. aeruginosa* freshly planted onto a smooth mica surface. This technology reveals the fine, three-dimensional detail of the bacterial surface structures.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The type IV pili found in *Neisseria* species, *Moraxella* species, *Vibrio cholerae*, *Legionella pneumophila*, *Salmonella enterica* serovar Typhi, enteropathogenic *E. coli*, and *Pseudomonas aeruginosa* often mediate adherence of organisms to target surfaces. Type IV pili tend to have a relatively conserved amino-terminal region and a more variable carboxyl-terminal region. For some species (e.g., *N. gonorrhoeae*, *Neisseria meningitidis*, and enteropathogenic *E. coli*), the pili are critical for attachment to mucosal epithelial cells. For others, such as *P. aeruginosa*, the pili may inhibit colonization; recent studies of *P. aeruginosa* colonization showed that,

in a bank of mutants in which all nonessential genes were interrupted, those unable to produce type IVa pili were actually better able to colonize the gastrointestinal and lung mucosa of mice. *V. cholerae* cells appear to use two different types of pili for intestinal colonization. Whereas interference with this stage of colonization would appear to be an effective antibacterial strategy, attempts to develop pilus-based vaccines against human diseases have not been highly successful to date.

Flagella are long appendages attached at one or both ends of the bacterial cell (*polar flagella*) or distributed over the entire cell surface (*peritrichous flagella*). Flagella, like pili, are composed of a polymerized or aggregated basic protein. In flagella, the protein subunits form a tight helical structure and vary serologically with the species. Spirochetes such as *T. pallidum* and *Borrelia burgdorferi* have axial filaments similar to flagella running down the long axis of the center of the cell, and they “swim” by rotation around these filaments. Some bacteria can glide over a surface in the absence of obvious motility structures.

Other bacterial structures involved in adherence to host tissues include staphylococcal and streptococcal proteins that bind to human extracellular matrix proteins such as fibrin, fibronectin, fibrinogen, laminin, and collagen. Fibronectin is a commonly used receptor for various pathogens; a particular amino acid sequence in fibronectin, Arg-Gly-Asp or RGD, is a conserved target used by bacteria to bind to host tissues. Binding of the *Staphylococcus aureus* surface protein clumping factor A (ClfA) to fibrinogen has been implicated in many aspects of pathogenesis. The conserved outer-core portion of the lipopolysaccharide (LPS) of *P. aeruginosa* mediates binding to the cystic fibrosis transmembrane conductance regulator (CFTR) on airway epithelial cells—an event that appears to be critical for normal host resistance to infection, initiating recruitment of polymorphonuclear neutrophils (PMNs) to the lung mucosa to kill the cells via opsonophagocytosis. A large number of microbial pathogens encompassing major gram-positive bacteria (staphylococci and streptococci), gram-negative bacteria (major enteric species and coccobacilli), fungi (*Candida*, *Fusobacterium*, *Aspergillus*), and even eukaryotic pathogens (*Trichomonas vaginalis* and *Plasmodium falciparum*) express a surface polysaccharide composed of β -1-6-linked-poly-*N*-acetyl-d-glucosamine (PNAG). One of its functions is to promote binding to synthetic materials used in catheters and other types of implanted devices. This polysaccharide may be a critical factor in the establishment of device-related infections by pathogens such as staphylococci and *E. coli*. High-powered imaging techniques (e.g., atomic force microscopy) have revealed that bacterial cells have a nonhomogeneous surface that is probably attributable to different concentrations of cell surface molecules, including microbial adhesins, at specific locations (Fig. 116-1, panels C and D).

Fungal Adhesins

Fungi produce adhesins that mediate colonization of epithelial surfaces, adhering particularly to structures like fibronectin, laminin, and collagen. The *Candida albicans* INT1 protein bears similarity to mammalian integrins that bind to extracellular matrix proteins. The agglutinin-like sequence (ALS) adhesins are large cell-surface glycoproteins mediating adherence of pathogenic *Candida* to host tissues. These adhesins possess a conserved three-domain structure composed of an N-terminus that mediates adherence to host tissue receptors, a central motif consisting of a number of repeats of a conserved sequence of 36 amino acids, and a C-terminal domain that varies in length and sequence and contains a glycosylphosphatidylinositol (GPI) anchor addition that allows the adhesins to bind to the fungal cell wall. Variability in the number of central domains characterizes different ALS proteins with specificity for different host receptors. The ALS adhesins are expressed under certain environmental conditions and are crucial for pathogenesis of fungal infections.

For several respiratory fungal pathogens, the inoculum is ingested by alveolar macrophages in which the fungal cells transform to pathogenic phenotypes. Like *C. albicans*, *Blastomyces dermatitidis* produces a 120-kDa surface protein, designated WI-1, that binds to CD11b/CD18 integrins as well as to CD14 on macrophages. An unidentified factor on *Histoplasma capsulatum* also mediates binding to the integrin surface proteins.

Eukaryotic Pathogen Adhesins

Eukaryotic parasites use complicated surface glycoproteins as adhesins, some of which are lectins (proteins that bind to specific carbohydrates on host cells). *Plasmodium vivax*, one of six *Plasmodium* species causing malaria, binds (via Duffy-binding protein) to the Duffy blood group carbohydrate antigen Fy on erythrocytes. *Entamoeba histolytica*, the third leading cause of death from parasitic diseases, expresses two proteins that bind to the disaccharide galactose/*N*-acetyl galactosamine. Children with mucosal IgA antibody to one of these lectins are resistant to reinfection with virulent *E. histolytica*. A major surface glycoprotein (gp63) of *Leishmania* promastigotes is needed for these parasites to enter human macrophages—the principal target cell of infection. This glycoprotein promotes complement binding but inhibits complement lytic activity, allowing the parasite to use complement receptors for entry into macrophages; gp63 also binds to fibronectin receptors on macrophages. As part of hepatic granuloma formation, *Schistosoma mansoni* expresses a carbohydrate epitope related to the Lewis X blood group antigen that promotes adherence of helminth eggs to vascular endothelial cells under inflammatory conditions.

Host Receptors

Host receptors are found both on target cells (such as epithelial cells lining mucosal surfaces) and within the mucus layer covering these cells. Microbial pathogens bind to a wide range of host receptors to establish infection (Table 116-1). Selective loss of host receptors for a pathogen may confer natural resistance to an otherwise susceptible population. For example, 70% of individuals in western Africa lack Fy antigens and are resistant to *P. vivax* infection. *S. enterica* serovar Typhi, the etiologic agent of typhoid fever, produces a pilus protein that binds to CFTR to enter the gastrointestinal submucosa after being ingested by enterocytes. As homozygous mutations in *CFTR* are the cause of the life-shortening disease cystic fibrosis, heterozygote carriers (e.g., 4–5% of individuals of European ancestry) may have had a selective advantage due to decreased susceptibility to typhoid fever.

Numerous virus–target cell interactions have been described, and it is now clear that different viruses can use similar host cell receptors for entry. The list of certain and likely host receptors for viral pathogens is long. Among the host membrane components that can serve as receptors for viruses are sialic acids, gangliosides, glycosaminoglycans, integrins and other members of the immunoglobulin superfamily, histocompatibility antigens, and regulators and receptors for complement components. An example of the effect of host receptors on the pathogenesis of infection has emerged from studies comparing the binding of avian influenza A virus subtype H5N1 with that of influenza A strains expressing the H1 hemagglutinin subtype. These subtypes are highly pathogenic and transmissible from human to human, and they bind to a receptor composed of two sugar molecules: sialic acid linked α -2-6 to galactose. This receptor is expressed at high levels in the human airway epithelium; when virus is shed from this surface, its transmission via coughing and aerosol droplets is facilitated. In contrast, the H5N1 avian influenza virus

binds to sialic acid linked α -2-3 to galactose, and this receptor is expressed at high levels on cells in the terminal bronchioles, including type II pneumocytes, alveolar macrophages, and nonciliated cuboidal epithelial cells. Infection at these sites is thought to underlie the high mortality rate associated with avian influenza but also the low interhuman transmissibility of this strain, which is not readily transported to the airways from which it can be expelled by coughing. Nonetheless, it has been shown that H5 hemagglutinins can acquire mutations leading to binding to α -2-6-linked sialic acids that increase their human transmissibility but retain their high level of lethality.

MICROBIAL GROWTH AFTER ENTRY

Once established on a mucosal or skin site, pathogenic microbes must replicate before causing full-blown infection and disease. Within cells, viral particles release their nucleic acids, which may be directly translated into viral proteins (positive-strand RNA viruses), transcribed from a negative strand of RNA into a complementary mRNA (negative-strand RNA viruses), or transcribed into a complementary strand of DNA (retroviruses). For DNA viruses, mRNA may be transcribed directly from viral DNA, either in the cell nucleus or in the cytoplasm. To grow, bacteria must acquire specific nutrients or synthesize them from precursors in host tissues. Many infectious processes are most often found in specific sites—e.g., H1 influenza in the respiratory mucosa, gonorrhea in the urogenital epithelium, and shigellosis in the gastrointestinal epithelium. While there are multiple reasons for this specificity, one important consideration is the ability of these pathogens to obtain the nutrients needed for growth and survival.

Temperature restrictions also play a role in limiting certain pathogens to specific tissues. Rhinoviruses, a cause of the common cold, grow best at 33°C and replicate in cooler nasal tissues but not in the lung. Leprosy lesions due to *Mycobacterium leprae* are found in and on relatively cool body sites. Fungal pathogens that infect the skin, hair follicles, and nails (dermatophyte infections) remain confined to the cooler, exterior, keratinous layer of the epithelium.

A topic of major interest is the ability of many bacterial, fungal, and protozoal species to grow in multicellular masses referred to as *biofilms*. These masses are biochemically and morphologically quite distinct from the free-living individual cells referred to as *planktonic cells*. Growth in biofilms leads to altered microbial metabolism, production of extracellular virulence factors, and decreased susceptibility to biocides, antimicrobial agents, and host defense molecules and cells. *P. aeruginosa* growing on the bronchial mucosa during chronic infection, staphylococci and other pathogens growing on implanted medical devices, and dental pathogens growing on tooth surfaces to form plaques represent several examples of microbial biofilm growth associated with human disease. Many other pathogens can form biofilms during in vitro growth. This mode of growth contributes to microbial virulence and induction of disease and can also be an important factor in microbial survival outside the host, promoting transmission to additional susceptible individuals.

AVOIDANCE OF INNATE HOST DEFENSES

Microbes have interacted with mucosal/epithelial surfaces since the emergence of multicellular organisms. Thus it is not surprising that multicellular hosts have a variety of innate surface defense mechanisms that can sense when pathogens are present and contribute to their elimination. The skin is acidic and bathed with fatty acids toxic to many microbes. Skin pathogens such as staphylococci must tolerate these adverse conditions. Mucosal surfaces are covered by a barrier composed of a thick mucus layer that entraps microbes and facilitates their transport out of the body by mucociliary clearance, coughing, and urination. Mucous secretions, saliva, and tears contain antibacterial factors such as lysozyme and antimicrobial peptides as well as antiviral factors such as interferons (IFNs). Gastric acidity and bile salts are inimical to the survival of many ingested organisms, and most mucosal surfaces—particularly the nasopharynx, vaginal tract, and gastrointestinal tract—contain a resident flora of commensal microbes that interfere with the ability of pathogens to colonize and infect a host. Major advances in the use of nucleic acid sequencing now allow extensive identification and characterization of the vast array of commensal organisms that have come to be referred to as the *microbiota*. In addition to its role in providing competition for mucosal colonization, acquisition of a normal microbiota is critical for proper development of the immune system, impacting maturation and differentiation of components of both the innate and acquired immune systems.

Pathogens that survive local antimicrobial factors must still contend with host endocytic, phagocytic, and inflammatory responses as well as with host genetic factors that determine the degree to which a pathogen can survive and grow. The growth of viral pathogens entering skin or mucosal epithelial cells can be limited by a variety of host genetic factors, including production of IFNs, modulation of receptors for viral entry, and age- and hormone-related susceptibility factors; by nutritional status; and even by personal habits such as smoking and exercise. The list of genes whose variants can affect host susceptibility and resistance to infection is rapidly expanding. A classic example is a 32-bp deletion in the gene for the HIV-1 co-receptor known as chemokine receptor 5 (CCR5), which, when present in the homozygous state, confers high-level resistance to HIV-1 infection. A now-famous case is that of the “Berlin Patient,” a man infected with HIV who received a hematopoietic stem-cell transplant from a donor homozygous for the 32-bp CCR5 deletion to treat acute myeloid leukemia. The apparent sterilizing cure of this patient’s HIV infection is likely due to his having only HIV-resistant T cells after the stem cell transplantation.

Encounters with Epithelial Cells

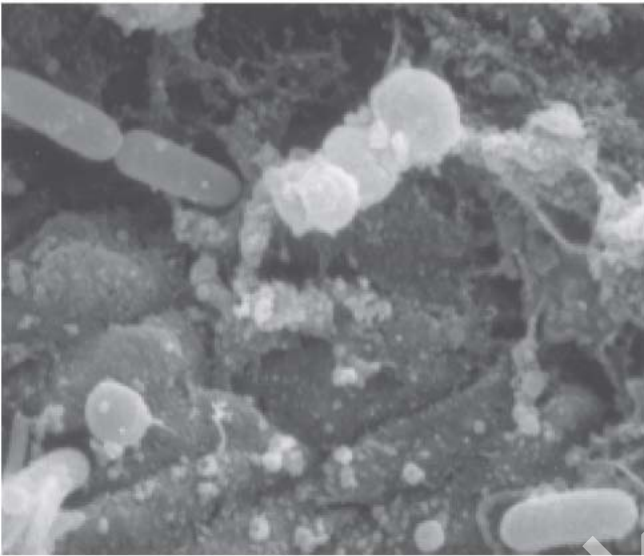
Over the past two decades, many pathogens have been shown to enter epithelial cells (Fig. 116-2) by using specialized surface structures that bind to receptors, with consequent internalization. However, the exact role and the importance of this process in infection and disease are not well defined for most of these pathogens. Microbial entry into host epithelial cells is seen as a path for translocation to adjacent or deeper tissues or as a route to a sanctuary site to avoid killing by professional phagocytes. Epithelial cell entry is a critical aspect of dysentery induction by *Shigella*.

FIGURE 116-2

Entry of bacteria into epithelial cells. A. Internalization of *Pseudomonas aeruginosa* by cultured airway epithelial cells expressing wild-type cystic fibrosis transmembrane conductance regulator, the cell receptor for bacterial ingestion. **B.** Entry of *P. aeruginosa* into murine tracheal epithelial cells after infection of mice by the intranasal route.



A



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Curiously, less virulent strains of many bacterial pathogens are more adept at entering epithelial cells than are more virulent strains; examples include pathogens that lack the surface polysaccharide capsule needed to cause serious disease. Thus, for *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Streptococcus agalactiae* (group B *Streptococcus*), *N. meningitidis*, and *S. pyogenes*, isogenic mutants or variants lacking capsules enter epithelial cells more easily than the wild-type, encapsulated parental forms that cause disseminated disease. These observations have led to the proposal that epithelial cell entry may be primarily a manifestation of host defense, resulting in bacterial clearance by both shedding of epithelial cells containing internalized bacteria and initiation of a protective and nonpathogenic inflammatory response. However, a possible consequence of this process could be the opening of a hole in the epithelium, potentially allowing uningested organisms to enter the submucosa. This scenario has been documented in murine *S. enterica* serovar Typhimurium infections and in experimental bladder infections caused by uropathogenic *E. coli*. In the latter system, bacterial pilus-mediated attachment to *uropilins* induces exfoliation of the cells with attached bacteria. Subsequently, infection is produced by residual bacterial cells that invade the superficial bladder epithelium, where they can grow intracellularly into biofilm-like masses encased in an extracellular polysaccharide-rich matrix and surrounded by uropilin. It is likely that at low bacterial inocula epithelial cell ingestion and subclinical inflammation are efficient means to eliminate pathogens, whereas at higher inocula a proportion of surviving bacterial cells enter host tissue through the damaged mucosal surface and multiply, producing disease. Alternatively, failure of the appropriate epithelial cell response to a pathogen may allow the organism to survive on a mucosal surface where, if it avoids other host defenses, it can grow and cause a local infection. Along these lines, as noted above, *P. aeruginosa* is taken into epithelial cells by CFTR, a protein missing or nonfunctional in most patients with severe cases of cystic fibrosis. The major clinical consequence of this disease is chronic airway-surface infection with *P. aeruginosa* in 80–90% of patients. The failure of airway epithelial cells to ingest and promote the removal of *P. aeruginosa* via a properly regulated inflammatory response has been proposed as a key component of the hypersusceptibility of these patients to chronic airway infection with this organism.

Encounters with Phagocytes

Phagocytosis and Inflammation

Phagocytosis of microbes is a major innate host defense that limits the growth and spread of pathogens. Phagocytes appear rapidly at sites of infection in conjunction with the initiation of inflammation. Ingestion of microbes by both tissue-fixed macrophages and migrating phagocytes probably accounts for the limited ability of most microbial agents to cause disease. A family of related molecules called *collectins*, *soluble defense collagens*, or *pattern-recognition molecules* are found in blood (mannose-binding lectins), lung (surfactant proteins A and D), and most likely other tissues and bind to carbohydrates on microbial surfaces to

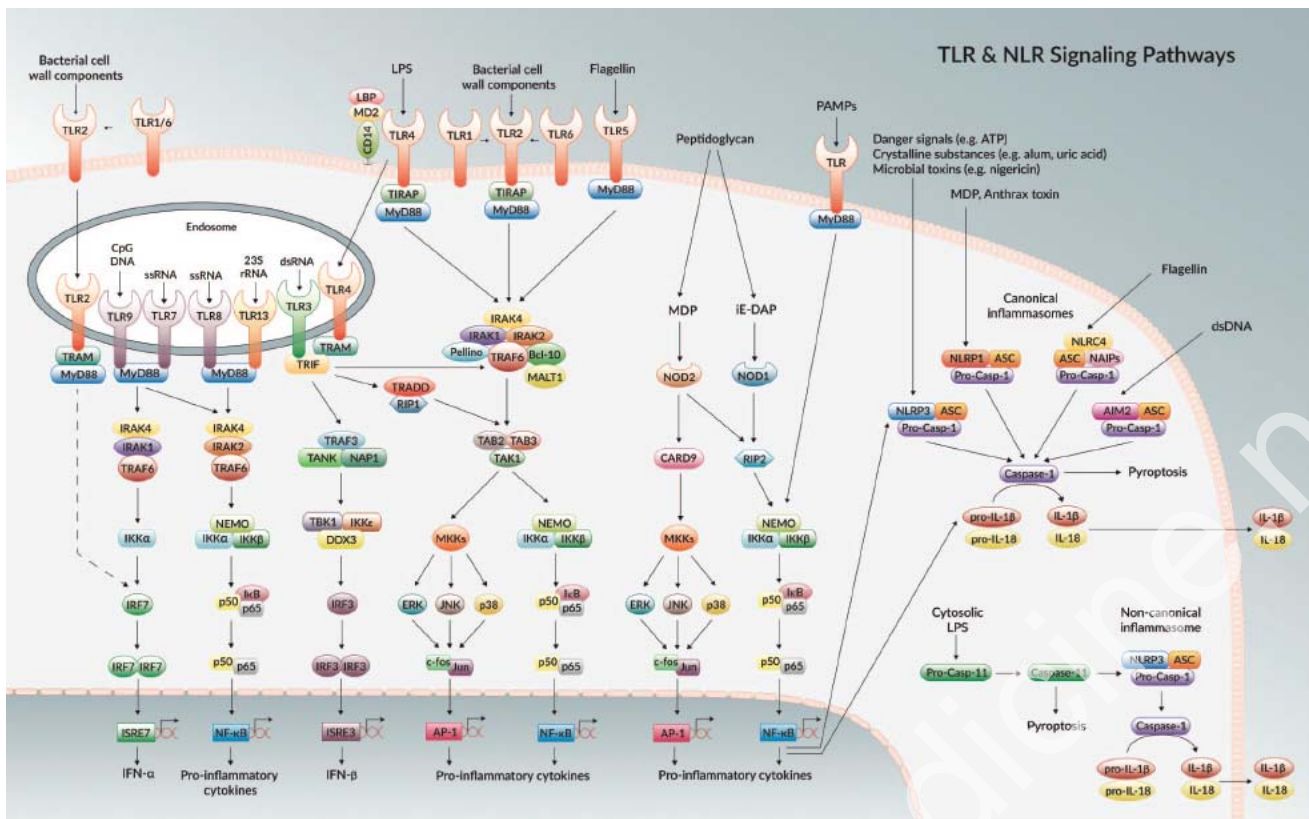
promote phagocyte clearance. Bacterial pathogens are ingested principally by PMNs, while eosinophils are frequently found at sites of infection by protozoan or multicellular parasites. Successful pathogens, by definition, must avoid being cleared by professional phagocytes. One of several antiphagocytic strategies employed by bacteria and by the fungal pathogen *Cryptococcus neoformans* is to elaborate large-molecular-weight surface polysaccharide antigens, often in the form of a capsule that coats the cell surface. Most pathogenic bacteria produce such antiphagocytic capsules. On occasion, proteins or polypeptides form capsule-like coatings for organisms such as group A streptococci and *Bacillus anthracis*.

An area of both intense interest and controversy is the role of the release of neutrophil extracellular traps (NETs) in protection against infection. NETs are composed of DNA and other intracellular components with antimicrobial properties, such as histones, myeloperoxidase, and elastase. NET release has been described as both a “suicidal” event, wherein, in response to stimuli, PMNs lyse and release NET components, and a “vital” event, wherein intracellular NET components are released but neutrophils remain viable and functional. Microbial particle size might regulate release of NETs, as has been reported for larger microbial structures like *C. albicans* hyphae or cellular aggregates. NET formation can also be pathologic as these networks are associated with damage to cells, thrombosis, inhibition of wound healing, and autoimmunity.

As activation of local phagocytes in tissues is a key step in initiating inflammation and migration of additional phagocytes into infected sites, much attention has been paid to microbial factors that initiate inflammation. These are usually conserved factors critical to the microbes' survival and are referred to as *pathogen-associated molecular patterns* (PAMPs). Cellular responses to microbial encounters with phagocytes are governed largely by the structure of the microbial PAMPs that elicit inflammation, and detailed knowledge of these structures of bacterial pathogens has contributed greatly to our understanding of molecular mechanisms of microbial pathogenesis mediated by activation of host cell molecules such as Toll-like receptors (TLRs; Fig. 116-3). One of the best-studied systems involves the interaction of LPS from gram-negative bacteria with the GPI-anchored membrane protein CD14 found on the surface of professional phagocytes, including migrating and tissue-fixed macrophages and PMNs. A soluble form of CD14 is also found in plasma and on mucosal surfaces. A plasma protein, LPS-binding protein, transfers LPS to membrane-bound CD14 on myeloid cells and promotes binding of LPS to soluble CD14. Soluble CD14/LPS/LPS-binding protein complexes bind to many cell types and may be internalized to initiate cellular responses to microbial pathogens. It has been shown that peptidoglycan and lipoteichoic acid from gram-positive bacteria and cell surface products of mycobacteria and spirochetes can interact with CD14 (Fig. 116-3). Additional molecules, such as MD-2, also participate in the recognition of bacterial activators of inflammation.

FIGURE 116-3

Toll-like receptor (TLR) and NOD-like receptor (NLR) signaling pathways. Microbial cell-surface constituents interact with TLRs, in some cases requiring additional factors such as MD2, which facilitates the response to lipopolysaccharide (LPS) via TLR4. Although microbial cell-surface constituents are depicted as interacting with the TLRs on the cell surface, TLRs contain extracellular leucine-rich domains that become localized to the lumen of the phagosome upon uptake of bacterial cells. The internalized TLRs can bind to microbial products. The TLRs are oligomerized, usually forming homodimers, and then bind to the general adapter protein MyD88 via the C-terminal Toll/interleukin 1 receptor (IL-1R) (TIR) domains, which also bind to TIRAP (TIR domain-containing adapter protein), a molecule that participates in the transduction of signals from TLRs 1, 2, 4, and 6. The MyD88/TIRAP complex activates signal-transducing molecules such as IRAK4 (IL-1R-associated kinase 4), which in turn activates IRAK1. This activation can be blocked by IRAKM and Toll interacting protein (TOLLIP). IRAK1 activates TRAF6 (tumor necrosis factor receptor-associated factor 6), TAK1 (transforming growth factor β -activating kinase 1), and TAB1/2 (TAK1-binding protein 1/2). This signaling complex associates with the ubiquitin-conjugating enzyme Ubc13 and the Ubc-like protein UEV1A to catalyze the formation of a polyubiquitin chain on TRAF6. Polyubiquitination of TRAF6 activates TAK1, which, along with TAB1/2 (a protein that binds to lysine residue 63 in polyubiquitin chains via a conserved zinc-finger domain), phosphorylates the inducible kinase complex: IKK α , - β , and - γ . IKK γ is also called NEMO [nuclear factor κ B (NF- κ B) essential modulator]. This large complex phosphorylates the inhibitory component of NF- κ B, I κ B α , resulting in release of I κ B α from NF- κ B. Phosphorylated (PP) I κ B is then ubiquitinated (ub) and degraded, and the two components of NF- κ B—p50 (or Rel) and p65—translocate to the nucleus, where they bind to regulatory transcriptional sites on target genes, many of which encode inflammatory proteins. In addition to inducing NF- κ B nuclear translocation, the TAK1/TAB1/2 complex activates MAP kinase transducers such as MKK 4/7 and MKK 3/6, an event that can lead to nuclear translocation of transcription factors such as AP1. TLR4 can also activate NF- κ B nuclear translocation via the MyD88-independent TRIF [TIR domain-containing, adapter-inducing interferon β (IFN- β)] and TRAM (TRIF-related adapter molecule) cofactors. Intracellular TLRs 3, 7, 8, and 9 also use MyD88 and TRIF to activate IFN response factors 3 and 7 (IRF3 and IRF7), which also function as transcriptional factors in the nucleus. The nucleotide-binding oligomerization domain-like receptor (NLR) family of proteins is involved in the regulation of innate immune responses. These proteins sense pathogen-associated molecular patterns (PAMPs) in the cytosol as well as the host-derived signals known as *damage-associated molecular patterns* (DAMPs). Certain NLRs induce the assembly of large caspase 1-activating complexes called *inflammasomes*. Activation of caspase 1 through autoproteolytic maturation leads to the processing and secretion of the proinflammatory cytokines interleukin 1 β (IL-1 β) and IL-18. So far, four inflammasomes have been identified and defined by the NLR protein they contain: the NLRP1/NALP1b inflammasome, the NLRP3/NALP3 inflammasome, the NLRP3/NALP3 inflammasome, and the AIM2 (absent in melanoma 2)-containing inflammasome. (Pathway diagram reproduced with permission from InvivoGen; www.invivogen.com/review-inflammasome.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

GPI-anchored receptors do not have intracellular signaling domains; therefore, it is the TLRs that transduce signals for cellular activation due to LPS binding. Binding of microbial factors to TLRs to activate signal transduction occurs in the phagosome—and not on the surface—of dendritic cells that have internalized the microbe. This binding is probably due to the release of the microbial surface factor from the cell in the environment of the phagosome, where the liberated factor can bind to its cognate TLRs. TLRs initiate cellular activation through a series of signal-transducing molecules (Fig. 116-3) that lead to nuclear translocation of the transcription factor nuclear factor κ B (NF- κ B), a master-switch for production of important inflammatory cytokines such as tumor necrosis factor α (TNF- α) and interleukin 1 (IL-1).

The initiation of inflammation can also occur with viral particles and other microbial products such as polysaccharides, enzymes, and toxins. Bacterial flagella activate inflammation by binding of a conserved sequence to TLR5. Some pathogens (e.g., *Campylobacter jejuni*, *Helicobacter pylori*, and *Bartonella bacilliformis*) make flagella that lack this sequence and do not bind to TLR5; thus efficient host responses to infection are prevented. Bacteria also produce a high proportion of DNA molecules with unmethylated CpG residues that activate inflammation through TLR9. TLR3 recognizes double-strand RNA, a pattern-recognition molecule produced by many viruses during their replicative cycle. TLR1 and TLR6 associate with TLR2 to promote recognition of acylated microbial proteins and peptides.

The myeloid differentiation factor 88 (MyD88) molecule and the Toll/IL-1R (TIR) domain-containing adapter protein (TIRAP) bind to the cytoplasmic domains of TLRs and also to receptors that are part of the IL-1 receptor families. Numerous studies have shown that MyD88/TIRAP-mediated transduction of signals from TLRs and other receptors is critical for innate resistance to infection, activating MAP kinases and NF- κ B and thereby leading to production of cytokines/chemokines. Mice lacking MyD88 are more susceptible than normal mice to infections with a broad range of pathogens. In one study, nine children homozygous for defective MyD88 genes had recurrent infections with *S. pneumoniae*, *S. aureus*, and *P. aeruginosa*—three bacterial species showing increased virulence in MyD88-deficient mice. The MyD88-deficient children seemed to have no greater susceptibility to other bacteria, viruses, fungi, or parasites. Another component of the MyD88-dependent signaling pathway is a molecule known as IL-1 receptor–associated kinase 4 (IRAK4). Individuals with a homozygous deficiency in genes encoding this protein are at increased risk for *S. pneumoniae* and *S. aureus* infections and, to some degree, *P. aeruginosa* infections as well.

Some TLRs (e.g., TLR3 and TLR4) can also activate signal transduction via a MyD88-independent pathway involving TIR domain-containing, adapter-inducing IFN- β (TRIF) and the TRIF-related adapter molecule (TRAM). Signaling through TRIF and TRAM activates the production of both NF- κ B-dependent cytokines/chemokines and type I IFNs. The type I IFNs bind to the IFN- α receptor composed of two protein chains, IFNAR1 and IFNAR2. Humans produce three type I IFNs: IFN- α , IFN- β , and IFN- γ . These molecules activate another class of proteins known as signal transducer and activator of transcription (STAT) complexes. The STAT factors are important in regulating immune system genes and thus play a critical role in responses to microbial infections.

Another intracellular complex of proteins found to be a major factor in the host cell response to infection is the *inflammasome* (Fig. 116-3), where inflammatory cytokines IL-1 and IL-18 are changed from their precursors to active forms by the *cysteine* protease caspase 1 and then secreted. The inflammasome is composed of proteins that are members of the nucleotide binding and oligomerization domain (NOD)-like receptor (NLR) family. Like the TLRs, NOD proteins within inflammasomes sense the presence of the conserved microbial factors either internalized from outside the cell to form canonical inflammasomes or released inside a cell (probably after microbial uptake) to directly activate pro-caspase-2. Recognition of these PAMPs by NLRs leads to caspase 1 or caspase 2 activation and to secretion of active IL-1 and IL-18. Studies of mice indicate that as many as four canonical inflammasomes with different components can be formed (Fig. 116-3). The

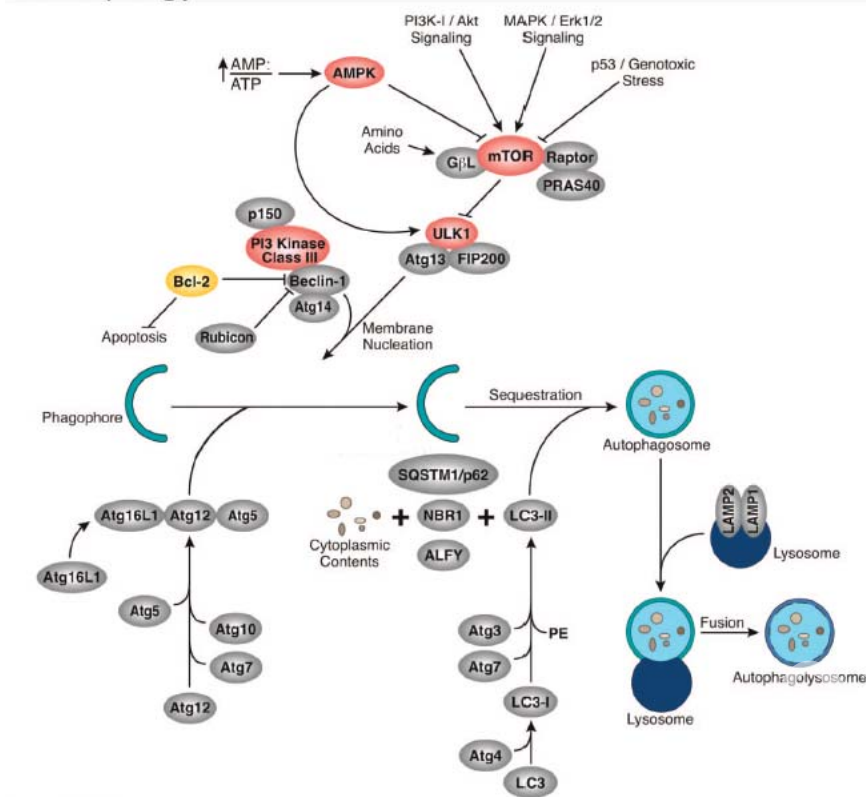
components depend on the type of stimulus driving inflammasome formation and activation. A fifth, noncanonical inflammasome responding to intracellular LPS also has been described (Fig. 116-3).

Some recent additions to the identified intracellular components responding to microbial infection are autophagy (Fig. 116-4), initially described as an intracellular process for degradation and recycling of cellular components for reuse, and several pathways leading to cell death, including apoptosis, RIPK1-dependent apoptosis, and necroptosis (Fig. 116-4). The latter three pathways are means by which cells undergo a death program in response to infection (notably, viral infection) and inflammation. Autophagy is an early defense mechanism mediated by caspases in response to pathogens wherein, after ingestion, microbes in either vacuoles or the cytoplasm are delivered to lysosomal compartments for degradation. Avoidance of this process is critical if pathogens are to cause disease. Pathogens can avoid autophagy by multiple mechanisms; examples include the inhibition of proteins within the autophagic vacuole by *Shigella*, the recruitment of host proteins to prevent autophagy of *Listeria monocytogenes*, and the inhibition of vacuole formation by *L. pneumophila*. In the death pathways, cell death to inhibit viral replication (Fig. 116-4) is mediated by a series of reactions commencing with TNF- α production and binding of this molecule to its receptor, TNFR1. In the two apoptotic pathways, the final steps are activation of effector caspases 3 and 7 and apoptotic cell death. In necroptosis, oligomers of the mixed-lineage kinase domain-like protein (MLKL) form and insert into the cell's plasma membrane; their insertion leads to lysis and release of damage-associated molecular patterns (DAMPs), resulting in protective innate immune responses. Finally, additional pathways of cell death are being described, including ferroptosis, oxytosis, parthanatos, pyronecrosis, and pyroptosis. The impact of these pathways on host-pathogen interactions is only beginning to be investigated.

FIGURE 116-4

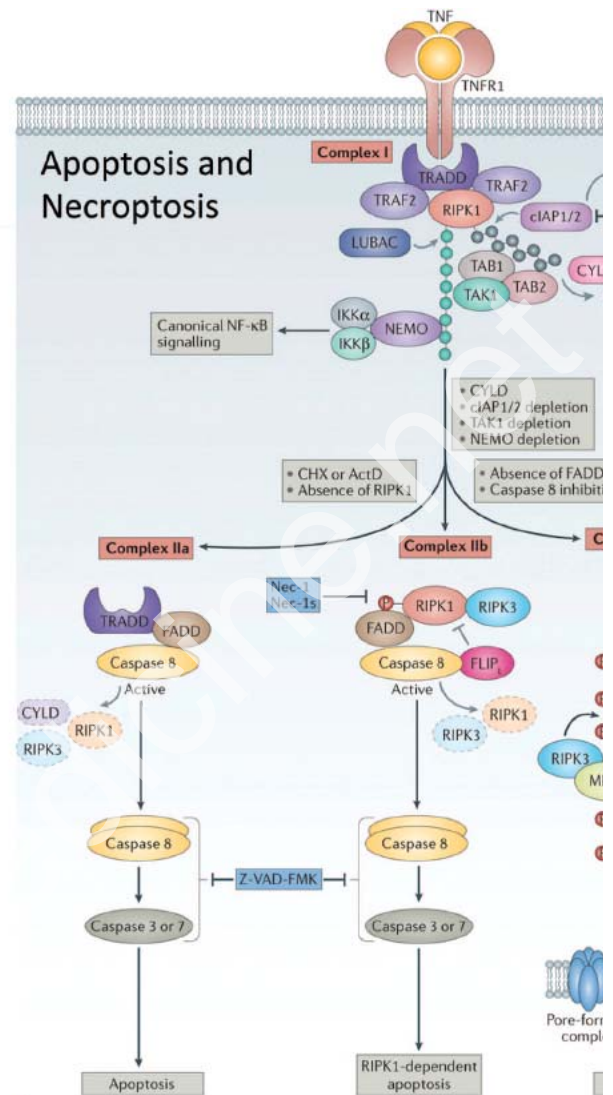
Autophagy, apoptosis, and necroptosis. A. Autophagy is a catabolic process that results in the autophagosomic-lysosomal degradation of bulk cytoplasmic contents, abnormal protein aggregates, and excess or damaged organelles. Autophagy is generally activated by conditions of nutrient deprivation but has also been associated with physiologic as well as pathologic processes such as development, differentiation, neurodegenerative disease, stress, infection, and cancer. The kinase mTOR is a critical regulator of autophagy induction, with activated mTOR (Akt and MAPK signaling) suppressing autophagy and negative regulation of mTOR (AMPK and p53 signaling) promoting it. Three related serine/threonine kinases—UNC-51-like kinases 1, 2, and 3 (ULK1, ULK2, ULK3), which play a role similar to that played by the yeast Atg1—act downstream of the mTOR complex. ULK1 and ULK2 form a large complex with the mammalian homolog of an autophagy-related (Atg) gene product (mAtg13) and the scaffold protein FIP200 (an ortholog of yeast Atg17). The class III PI3K complex, containing hVps34, beclin-1 (a mammalian homolog of yeast Atg6), p150 (a mammalian homolog of yeast Vps15), and Atg14-like protein (Atg14L or Borkor) or ultraviolet irradiation resistance-associated gene (UVRAG), is required for the induction of autophagy. The Atg genes control autophagosome formation through Atg12-Atg5 and LC3-II (Atg8-II) complexes. Atg12 is conjugated to Atg5 in a ubiquitin-like reaction that requires Atg7 and Atg10 (E1- and E2-like enzymes, respectively). The Atg12-Atg5 conjugate then interacts noncovalently with Atg16 to form a large complex. LC3/Atg8 is cleaved at its C-terminus by Atg4 protease to generate the cytosolic LC3-I. LC3-I is conjugated to phosphatidylethanolamine (PE), also in a ubiquitin-like reaction that requires Atg7 and Atg3 (E1- and E2-like enzymes, respectively). The lipidated form of LC3, known as LC3-II, is attached to the autophagosome membrane. Autophagy and apoptosis are connected both positively and negatively, and extensive crosstalk exists between the two processes. During nutrient deficiency, autophagy functions as a pro-survival mechanism; however, excessive autophagy may lead to cell death, a process morphologically distinct from apoptosis. Several pro-apoptotic signals, such as tumor necrosis factor (TNF), TNF-related apoptosis-inducing ligand (TRAIL), and FAS-associated death domain (FADD), also induce autophagy. Moreover, Bcl-2 inhibits beclin-1-dependent autophagy, thereby functioning as both a pro-survival and an anti-autophagic regulator. Mitophagy is a selective autophagic process specifically designed for the removal of damaged or unneeded mitochondria from a cell. Upon mitochondrial damage, the protein PINK, which is continually degraded in the healthy state through the action of PARL, is stabilized and recruits the E3 ligase Parkin to initiate mitophagy. Polyubiquitination of mitochondrial membrane proteins by Parkin results in the recruitment of autophagy adaptor proteins SQSTM1/p62, NBR1, and Ambra1, which bind to LC3 via their LC3-interacting region (LIR). In addition, BNIP3 and BNIP3L/NIX, which also contain LIRs, directly recruit autophagic machinery by a ubiquitin-independent mechanism to induce autophagosome formation in certain cell types. [Reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).] **B.** Apoptosis and necroptosis are initiated by the binding of TNF to its cognate receptor TNFR1, triggering the assembly of complex I, which comprises TNFR1, TNFR1-associated death domain (TRADD), receptor-interacting serine/threonine-protein kinase 1 (RIPK1), TNFR-associated factor 2 (TRAF2), cellular inhibitor of apoptosis protein 1/2 (cIAP1/2), and linear ubiquitin chain assembly complex (LUBAC). Complex I provides the platform for Lys63-linked ubiquitylation (gray circles) or linear ubiquitylation (green circles) of RIPK1 by cIAP1/2 and LUBAC, respectively. This ubiquitylation is implicated in the decision between NF- κ B or survival signaling and cell death signaling. Ubiquitylation leads to the recruitment of other factors, such as transforming growth factor β -activated kinase (TAK1), TAK1-binding protein 1 (TAB1), TAB2, NF- κ B essential modulator (NEMO), and the inhibitor of the NF- κ B kinase- α (IKK α)–IKK β complex; this recruitment usually triggers canonical NF- κ B signaling. In the presence of the translational inhibitor cycloheximide (CHX), TNF stimulation leads to the formation of cytosolic complex IIa, in which RIPK1 disappears, whereas interaction of TRADD and FADD leads to the activation of caspase 8 and effector caspases (e.g., caspase 3 and caspase 7) and to apoptotic cell death. With cIAP1/2 inhibitors [second mitochondria-derived activator of caspase (SMAC) mimetics], knockdown of cIAPs, or inhibition or depletion of TAK1 or NEMO, complex IIb is formed. Complex IIb consists of RIPK1, RIPK3, FADD, and caspase 8 and favors RIPK1-kinase activity–dependent apoptosis. Heteromeric pro-caspase 8–FLICE-like inhibitory protein long isoform (FLIP_L) inhibits necroptosis, probably by cleaving RIPK1, RIPK3, and cylindromatosis (CYLD). With sufficient expression of RIPK3 and concomitant inhibition or reduced expression of pro-caspase 8 and FLIP_L, complex IIc (also known as the necrosome) is formed. The formation of complex IIc entails the association of RIPK1 and RIPK3 followed by a series of auto- and transphosphorylation events of RIPK1 and RIPK3. Activated RIPK3 phosphorylates and recruits mixed-lineage kinase domain-like protein (MLKL), eventually leading to the formation of a supramolecular protein complex at the plasma membrane and necroptosis. SMAC mimetics are being developed to impair survival signaling and to sensitize cells to the triggering of cell death in the context of tumor treatment. Z-VAD-FMK is a bona fide pan-caspase inhibitor. Necrostatin-1 (Nec-1) and the more specific Nec-1s, Cpd27, and (more recently) PN10—a hybrid molecule consisting of Nec-1s and ponatinib—all inhibit the kinase activity of RIPK1 and thus inhibit necroptotic signaling. Additional necroptosis inhibitors include the RIPK3 inhibitors GSK'840, GSK'843 and GSK'872 as well as the MLKL inhibitors necrosulfonamide (NSA) and compound 1. However, the specificity of these MLKL inhibitors is not restricted to inhibition of MLKL, and their efficacy is probably also due to effects on other steps in the necroptosis pathway. ActD, actinomycin D. (Reprinted by permission from Macmillan Publishers Ltd: M Conrad et al: Regulated necrosis: Disease relevance and therapeutic opportunities. Nature Reviews Drug Discovery 15:348, copyright 2016.)

Autophagy



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.



B

Additional Interactions of Microbial Pathogens and Phagocytes

Other ways that microbial pathogens avoid destruction by phagocytes include production of factors that are toxic to these cells or that interfere with their chemotactic and ingestion function. Hemolysins, leukocidins, and the like are microbial proteins that can kill phagocytes. *S. aureus* elaborates a family of bi-component leukocidins that bind to host receptors such as the HIV co-receptor CCR5, which is also a receptor for the LukE/D toxin, and the receptor of the C5a component of activated complement used by LukF/S, also known as the *Panton-Valentine leukocidin*. The cytolytic staphylococcal α hemolysin binds to the disintegrin and metalloprotease 10 (ADAM-10) protein expressed on a variety of cells and also activates the NLRP3 inflammasome in monocytic cells, with consequent production of inflammatory cytokines as well as cell death. Streptolysin O made by *S. pyogenes* binds to cholesterol in phagocyte membranes and initiates a process of internal degranulation, with the release of normally granule-sequestered toxic components into the phagocyte's cytoplasm. *E. histolytica*, an intestinal protozoan that causes amebic dysentery, can disrupt phagocyte membranes after direct contact via the release of protozoal phospholipase A and pore-forming peptides.

Microbial Survival Inside Phagocytes

Many important microbial pathogens use a variety of strategies to survive inside phagocytes (particularly macrophages) after ingestion. Inhibition of fusion of the phagocytic vacuole (the phagosome) containing the initially ingested microbe with the lysosomal granules containing antimicrobial substances (the lysosome) allows *M. tuberculosis*, *S. enterica* serovar Typhi, and *Toxoplasma gondii* to survive inside macrophages. Some organisms, such as *L. monocytogenes*, escape into the phagocyte's cytoplasm to grow and eventually spread to other cells. Resistance to killing within the macrophage and subsequent growth are critical to successful infection by herpes-type viruses, measles virus, poxviruses, *Salmonella*, *Yersinia*, *Legionella*, *Mycobacterium*, *Trypanosoma*, *Nocardia*, *Histoplasma*, *Toxoplasma*, and *Rickettsia*. *Salmonella* species use a master regulatory system—in which the *PhoP/PhoQ* genes control other genes—to enter and survive within cells, with intracellular survival entailing structural changes in the cell envelope LPS.

TISSUE INVASION AND TISSUE TROPISM

Tissue Invasion

Most viral pathogens cause disease by growth at skin or mucosal entry sites, but some pathogens spread from the initial site to deeper tissues. Virus can spread via the nerves (rabies virus) or plasma (picornaviruses) or within migratory blood cells (poliovirus, Epstein-Barr virus, and many others). Specific viral genes determine where and how individual viral strains can spread.

Bacteria may invade deeper layers of mucosal tissue via intracellular uptake by epithelial cells, traversal of epithelial cell junctions, or penetration through denuded epithelial surfaces. Among virulent *Shigella* strains and invasive *E. coli*, outer-membrane proteins are critical to epithelial cell invasion and bacterial multiplication. *Neisseria* and *Haemophilus* species penetrate mucosal cells by poorly understood mechanisms before dissemination into the bloodstream. Staphylococci and streptococci elaborate a variety of extracellular enzymes, such as hyaluronidase, lipases, nucleases, and hemolysins, that are probably important in breaking down cellular and matrix structures and allowing the bacteria access to deeper tissues and blood. Staphylococcal α hemolysin binding to ADAM-10 leads to endothelial cell damage and disruption of vascular barrier function, events that are probably critical for systemic spread of *S. aureus* from an initial infectious site. Organisms that colonize the gastrointestinal tract can often translocate through the mucosa into the blood and, under circumstances in which host defenses are inadequate, cause bacteremia. *Yersinia enterocolitica* can invade the mucosa through the activity of the invasin protein. The complex milieu of basement membrane-containing structures such as laminin and collagen that anchor epithelial cells to mucosal surfaces must often be breached. A family of microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) can attach bacteria to the extracellular matrix and, along with proteases that degrade the basement proteins as well as surface-bound plasminogen and matrix metalloproteinases recruited from the host, permit breaching of this structure. Some bacteria (e.g., *Brucella*) can be carried from a mucosal site to a distant site by phagocytic cells that ingest but fail to kill the bacteria.

Fungal pathogens almost always take advantage of host immunocompromise to spread hematogenously to deeper tissues. The AIDS epidemic has resoundingly illustrated this principle: the immunodeficiency of many HIV-infected patients permits the development of life-threatening fungal infections of the lung, blood, and brain. Other than the capsule of *C. neoformans*, specific fungal antigens involved in tissue invasion are not well characterized. Both fungal pathogens and protozoal pathogens (e.g., *Plasmodium* species and *E. histolytica*) undergo morphologic changes to spread within a host. *C. albicans* undertakes a yeast-hyphal transformation wherein the hyphal forms are found where the fungus is infiltrating the mucosal barrier of tissues, while the yeast form grows on epithelial cell surfaces as well as the tips of hyphae that have infiltrated tissues. Malarial parasites grow in liver cells as merozoites and are released into the blood to invade erythrocytes and become trophozoites. *E. histolytica* is found as both a cyst and a trophozoite in the intestinal lumen, through which this pathogen enters the host, but only the trophozoite form can spread systemically to cause amebic liver abscesses. Other protozoal pathogens, such as *T. gondii*, *Giardia lamblia*, and *Cryptosporidium*, also undergo extensive morphologic changes after initial infection to spread to other tissues.

Tissue Tropism

While it is well known that certain microbes cause disease by infecting specific tissues, the molecular basis for tissue tropism is understood somewhat better for viral pathogens than for other infectious agents. Specific receptor–ligand interactions clearly underlie the ability of certain viruses to enter cells within tissues and disrupt normal tissue function, but the mere presence of a receptor for a virus on a target tissue is not sufficient for tissue tropism. Factors in the cell, route of viral entry, viral capacity to penetrate into cells, viral genetic elements that regulate gene expression, and pathways of viral spread in a tissue all affect tissue tropism. Some viral genes are best transcribed in specific target cells, such as hepatitis B genes in liver cells and Epstein-Barr virus genes in B lymphocytes. The route of inoculation of poliovirus determines its neurotropism, although the molecular basis for this association is not understood.

Compared with viral tissue tropism, the tissue tropism of bacterial and parasitic infections has not been as clearly elucidated, but studies of *Neisseria* species have provided insights. Both *N. gonorrhoeae*, which colonizes and infects the human genital tract, and *N. meningitidis*, which principally colonizes the human oropharynx but can spread to the brain, produce type IV pili (Tfp) that mediate adherence to host tissues. In the case of *N. gonorrhoeae*, the Tfp bind to a glucosamine-galactose-containing adhesin on the surface of cervical and urethral cells; in the case of *N. meningitidis*, the Tfp bind to cells in the human meninges in order to cross the blood–brain barrier. *N. gonorrhoeae* can use cytidine monophosphate *N*-acetylneuraminic acid from host tissues to add *N*-acetylneuraminic acid (sialic acid) to its lipooligosaccharide O side chain, and this alteration makes the organism resistant to host defenses. Lactate, present at high levels on genital mucosal surfaces, stimulates sialylation of gonococcal lipooligosaccharide. Bacteria with sialic acid sugars in their capsules, such as *N. meningitidis*, *E. coli* K1, and group B streptococci, have a propensity to cause meningitis, but this generalization has many exceptions. For example, all recognized serotypes of group B streptococci contain sialic acid in their capsules, but only one serotype (III) is responsible for most cases of group B streptococcal meningitis. Moreover, both *H. influenzae* and *S. pneumoniae* can readily cause meningitis, but these organisms do not have sialic acid in their capsules.

TISSUE DAMAGE AND DISEASE

Disease is a complex phenomenon resulting from tissue invasion and destruction, toxin elaboration, and host response. Viruses cause much of their damage by exerting a cytopathic effect on host cells and inhibiting host defenses. The growth of bacterial, fungal, and protozoal parasites in tissue, which may or may not be accompanied by toxin elaboration, can also compromise tissue function and lead to disease. For some bacterial and possibly some fungal pathogens, toxin production is one of the best-characterized molecular mechanisms of pathogenesis, while host factors such as IL-1, TNF- α , kinins, inflammatory proteins, products of complement activation, and mediators derived from arachidonic acid metabolites (leukotrienes) and cellular degranulation (histamines) readily contribute to the severity of disease.

Viral Disease

Viral pathogens inhibit host immune responses by a variety of mechanisms—e.g., by decreasing production of major histocompatibility complex molecules (adenovirus E3 protein), diminishing cytotoxic T cell recognition of virus-infected cells (Epstein-Barr virus nuclear antigen 1 and cytomegalovirus intermediate-early protein), producing virus-encoded complement receptor proteins that protect infected cells from complement-mediated lysis (herpesvirus and vaccinia virus),

making proteins that interfere with the action of IFN (influenza virus and poxvirus), and elaborating superantigen-like proteins (mouse mammary tumor virus and related retroviruses and the rabies nucleocapsid). Superantigens activate large populations of T cells that express particular subsets of the T cell receptor β protein, causing massive cytokine release and subsequent host reactions. Dengue virus is the most common insect-transmitted virus in the world, causing symptoms ranging from none to serious systemic illness or “break-bone” fever (severe fever and pain). Along with the recent epidemic of the related flavivirus Zika virus, disruptions to host innate immunity and inhibition of programmed cell death that allows continued viral replication underlie the ability of these viruses to cause infections. Infections of pregnant women with Zika virus can result in viral crossing of the placenta; viral entry into and growth in fetal brain tissues result in the birth of neonates with microcephaly. Viruses also produce peptide growth factors for host cells, which disrupt normal cellular growth, proliferation, and differentiation. In addition, viral factors can bind to and interfere with the function of host receptors for signaling molecules. Modulation of cytokine production during viral infection can stimulate viral growth inside cells with receptors for the cytokine, and virus-encoded cytokine homologues (e.g., the Epstein-Barr virus BCRF1 protein, which is highly homologous to the immunoinhibitory IL-10 molecule) can prevent immune-mediated clearance of viral particles. Viruses cause disease in neural cells by interfering with levels of neurotransmitters without necessarily destroying the cells, or they may induce either programmed cell death (apoptosis) to destroy tissues or inhibitors of apoptosis to allow prolonged viral infection of cells. For infection to spread, many viruses must be released from cells. The HIV protein U (Vpu) facilitates virus release, a process that is specific to certain cells. Mammalian cells produce a restriction factor that inhibits release of some viruses; for HIV, this factor is designated BST-2 (bone marrow stromal antigen 2)/HM1.24/CD317, or tetherin. Vpu of HIV interacts with tetherin, allowing release of infectious virus. Overall, virus-induced disruption of normal cellular and tissue function promotes clinical disease.

Bacterial Toxins

Among the first infectious diseases to be understood were those due to toxin-elaborating bacteria. Diphtheria, botulism, and tetanus toxins are responsible for the diseases associated with local infections due to *Corynebacterium diphtheriae*, *Clostridium botulinum*, and *Clostridium tetani*, respectively. *C. difficile* is an anaerobic gram-positive organism that elaborates two toxins, A and B, responsible for disruption of the intestinal mucosa when its numbers expand in the intestine, leading to antibiotic-associated diarrhea and potentially pseudomembranous colitis. A clinical trial evaluating prevention of recurrence of *C. difficile* infection by monoclonal antibodies to toxins A and B showed that antibody to toxin B, but not toxin A, had a significant impact. Enterotoxins produced by *E. coli*, *Salmonella*, *Shigella*, staphylococci, and *V. cholerae* contribute to diarrheal disease caused by these organisms. Staphylococci, streptococci, *P. aeruginosa*, and *Bordetella* elaborate various toxins that cause or contribute to disease, including toxic shock syndrome toxin 1; erythrogenic toxin; exotoxins A, S, T, and U; and pertussis toxin. A number of these toxins (e.g., cholera toxin, diphtheria toxin, pertussis toxin, *E. coli* heat-labile toxin, and *P. aeruginosa* exotoxin) have adenosine diphosphate ribosyl transferase activity; i.e., the toxins enzymatically catalyze the transfer of the adenosine diphosphate ribosyl portion of nicotinamide adenine diphosphate to target proteins and inactivate them. The staphylococcal enterotoxins, toxic shock syndrome toxin 1, and the streptococcal pyogenic exotoxins behave as superantigens, stimulating certain T cells to proliferate without processing of the protein toxin by antigen-presenting cells. Part of this process involves stimulation of the antigen-presenting cells to produce IL-1 and TNF- α , which have been implicated in many clinical features of diseases like toxic shock syndrome and scarlet fever. A number of gram-negative pathogens (*Salmonella*, *Yersinia*, and *P. aeruginosa*) can inject toxins directly into host target cells by means of a complex set of proteins referred to as the *type III secretion system*. Loss or inactivation of this virulence system usually greatly reduces the capacity of a bacterial pathogen to cause disease.

Endotoxin

The lipid A portion of LPS in some gram-negative bacteria has potent biologic activities that cause many of the clinical manifestations of gram-negative bacterial sepsis, including fever, muscle proteolysis, uncontrolled intravascular coagulation, and shock. The effects of lipid A appear to be mediated by the production of potent cytokines due to LPS binding to CD14 and signal transduction via TLRs, particularly TLR4. Cytokines exhibit potent hypothermic activity through effects on the hypothalamus; they also increase vascular permeability, alter the activity of endothelial cells, and induce endothelial-cell procoagulant activity. Numerous therapeutic strategies aimed at neutralizing the effects of endotoxin are under investigation, and while studies with laboratory animals have been promising, they have not yet translated into positive results for human septic shock.

Invasion

Many diseases are caused primarily by pathogens growing to high levels in tissues. Pneumococcal pneumonia is mostly attributable to the growth of *S. pneumoniae* in the lung and the attendant host inflammatory response, although specific factors that enhance this process (e.g., pneumolysin) may be responsible for some of the pathogenic properties of the pneumococcus. Disease that follows bacteremia and invasion of the meninges by meningitis-producing bacteria such as *N. meningitidis*, *H. influenzae*, *E. coli* K1, and group B streptococci appears to be due solely to the ability of these organisms to gain access to these tissues, multiply in them, and provoke cytokine production leading to tissue-damaging host inflammation.

Specific molecular mechanisms accounting for tissue invasion by fungal and protozoal pathogens are less well described. Except for studies pointing to factors like capsule and melanin production by *C. neoformans*, and possibly levels of cell wall glucans in some pathogenic fungi, the molecular basis for fungal invasiveness is not well defined. Melanin has been shown to protect the fungal cell against death caused by phagocyte factors such as nitric oxide, superoxide, and hypochlorite. Morphogenic variation and production of proteases (e.g., the *Candida* aspartyl proteinase) have been implicated in fungal invasion of host tissues.

If pathogens are to invade host tissues (particularly the blood), they must avoid the major host defenses represented by complement and phagocytic cells. Bacteria most often avoid these defenses through their surface polysaccharides—either capsular polysaccharides or long O-side-chain antigens characteristic of the smooth LPS of gram-negative bacteria. These molecules can prevent the activation and/or deposition of complement opsonins or can limit the access of phagocytic cells with receptors for complement opsonins to these molecules when they are deposited on the bacterial surface below the capsular layer. Another potential mechanism of microbial virulence is the ability of some organisms to present the capsule as an apparent self antigen through molecular mimicry. For example, the polysialic acid capsule of group B *N. meningitidis* is chemically identical to an oligosaccharide found on human brain cells.

Immunochemical studies of capsular polysaccharides have led to an appreciation of the tremendous chemical diversity that can result from the linking of a few monosaccharides. For example, three hexoses can link up in more than 300 different, potentially serologically distinct ways, while three amino acids have only six possible peptide combinations. Capsular polysaccharides have been used as effective vaccines against meningococcal meningitis as well as against pneumococcal and *H. influenzae* infections and may prove to be of value as vaccines against any organism that expresses a nontoxic, immunogenic capsular polysaccharide. In addition, most encapsulated pathogens become virtually avirulent when capsule production is interrupted by genetic manipulation; this observation emphasizes the importance of this structure in pathogenesis. A capsule-like surface polysaccharide, PNAG, has been found as a conserved structure shared by many microbes but generally is a poor target for antibody-mediated immunity because of the propensity of most humans and animals, all of which are colonized by PNAG-producing microbes, to produce a nonprotective type of antibody. Altering the structure of PNAG by removing the acetate substituents on the *N*-acetyl glucosamine monomers leads to an immunogenic form, deacetylated PNAG, that has been reported to induce antibodies that are protective in animals against diverse microbial pathogens.

Host Response

The inflammatory response of the host is critical for interruption and resolution of the infectious process but also is often responsible for the signs and symptoms of disease. Infection promotes a complex series of host responses involving the complement, kinin, and coagulation pathways. The production of cytokines such as IL-1, IL-17, IL-18, TNF- α , IFN- γ , and other factors regulated in part by the NF- κ B transcription factor leads to fever, muscle proteolysis, and other effects. An inability to kill or contain the microbe usually results in further damage due to the progression of inflammation and infection. For example, in many chronic infections, degranulation of host inflammatory cells leads to release of host proteases, elastases, histamines, and other toxic substances that can degrade host tissues. Chronic inflammation in any tissue can lead to the destruction of that tissue and to clinical disease associated with loss of organ function, such as sterility from pelvic inflammatory disease caused by chronic infection with *N. gonorrhoeae*.

The nature of the host response elicited by the pathogen often determines the pathology of a particular infection. Local inflammation produces local tissue damage, while systemic inflammation, such as that seen during sepsis, can result in the signs and symptoms of septic shock. The severity of septic shock is associated with the degree of production of host effectors. Disease due to intracellular parasitism results from the formation of granulomas wherein the host attempts to wall off the parasite inside a fibrotic lesion surrounded by fused epithelial cells that make up so-called multinucleated giant cells. A number of pathogens, particularly anaerobic bacteria, staphylococci, and streptococci, provoke the formation of an abscess, probably because of the presence of zwitterionic surface polysaccharides such as the capsular polysaccharide of *Bacteroides fragilis*. The outcome of an infection depends on the balance between an effective host response that eliminates a pathogen and an excessive inflammatory response that is associated with an inability to eliminate a pathogen and with the resultant tissue damage that leads to disease.

TRANSMISSION TO NEW HOSTS

As part of the pathogenic process, most microbes are shed from the host, often in a form infectious for susceptible individuals. However, the rate of transmissibility may not necessarily be high, even if the disease is severe in the infected individual, as these traits are not linked. Most pathogens exit via the same route by which they entered: respiratory pathogens by aerosols from sneezing or coughing or through salivary spread, gastrointestinal pathogens by fecal–oral spread, sexually transmitted diseases by venereal spread, and vector-borne organisms by either direct contact with the vector through a blood meal or indirect contact with organisms shed into environmental sources such as water. Microbial factors that specifically promote transmission are not well characterized. Respiratory shedding is facilitated by overproduction of mucous secretions, with consequently enhanced sneezing and coughing. Diarrheal toxins such as cholera toxin, *E. coli* heat-labile toxins, and *Shigella* toxins probably facilitate fecal–oral spread of microbial cells in the high volumes of diarrheal fluid produced during infection. The ability to produce phenotypic variants that resist hostile environmental factors (e.g., the highly resistant cysts of *E. histolytica* shed in feces) represents another mechanism of pathogenesis relevant to transmission. Blood parasites such as *Plasmodium* species change phenotype after ingestion by a mosquito—a prerequisite for the continued transmission of this pathogen. Venereally transmitted pathogens may undergo phenotypic variation due to the production of specific factors to facilitate transmission, but shedding of these pathogens into the environment does not result in the formation of infectious foci.

SUMMARY

In summary, the molecular mechanisms used by pathogens to colonize, invade, infect, and disrupt the host are numerous and diverse. Each phase of the infectious process involves a variety of microbial and host factors interacting in a manner that can result in disease. Recognition of the coordinated genetic regulation of virulence factor elaboration when organisms move from their natural environment into the mammalian host emphasizes the complex nature of the host–parasite interaction. Fortunately, the need for diverse factors in successful infection and disease implies that a variety of therapeutic strategies may be developed to interrupt this process and thereby prevent and treat microbial infections.

FURTHER READING

Bäumler AJ, Sperandio V: Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535:85, 2016. [[PubMed: 27383983](#)]

Brunke S et al: Virulence factors in fungal pathogens of man. *Curr Opin Microbiol* 32:89, 2016. [[PubMed: 27257746](#)]

Fuchs Y, Steller H: Live to die another way: Modes of programmed cell death and the signals emanating from dying cells. *Nat Rev Mol Cell Bio* 16:329, 2016.

Honda K, Littman DR: The microbiota in adaptive immune homeostasis and disease. *Nature* 535:75, 2016. [[PubMed: 27383982](#)]

Olagnier D et al: Dengue virus immunopathogenesis: Lessons applicable to the emergence of Zika virus. *J Mol Biol* 428:3429, 2016. [[PubMed: 27130436](#)]

3/14/2018

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 117: Approach to the Acutely Ill Infected Febrile Patient

Tamar F. Barlam; Dennis L. Kasper

INTRODUCTION

The physician treating the acutely ill febrile patient must be able to recognize infections that require emergent attention. If such infections are not adequately evaluated and treated at initial presentation, the opportunity to alter an adverse outcome may be lost. In this chapter, the clinical presentations of and approach to patients with relatively common infectious disease emergencies are discussed. These infectious processes and their treatments are discussed in detail in other chapters.

APPROACH TO THE PATIENT

APPROACH TO THE PATIENT

Acute Febrile Illness

Before the history is elicited and a physical examination is performed, an immediate assessment of the patient's general appearance can yield valuable information. The perceptive physician's subjective sense that a patient is septic or toxic often proves accurate. Visible agitation or anxiety in a febrile patient can be a harbinger of critical illness.

HISTORY

Presenting symptoms are frequently nonspecific. Detailed questions should be asked about the onset and duration of symptoms and about changes in severity or rate of progression over time. Host factors and comorbid conditions may increase the risk of infection with certain organisms or of a more fulminant course than is usually seen. Lack of splenic function, alcoholism with significant liver disease, IV drug use, HIV infection, diabetes, malignancy, organ transplantation, and chemotherapy all predispose to specific infections and frequently to increased severity. The patient should be questioned about factors that might help identify a nidus for invasive infection, such as recent upper respiratory tract infections, influenza, or varicella; prior trauma; disruption of cutaneous barriers due to lacerations, burns, surgery, body piercing, or decubiti; and the presence of foreign bodies, such as nasal packing after rhinoplasty, tampons, or prosthetic joints. Travel, contact with pets or other animals, or activities that might result in tick or mosquito exposure can lead to diagnoses that would not otherwise be considered. Recent dietary intake, medication use, social

or occupational contact with ill individuals, vaccination history, recent sexual contacts, and menstrual history may be relevant. Pregnancy might increase the risk and severity of some illnesses, such as influenza, or increase the risk of significant morbidity for the fetus, as in *Listeria* or Zika virus infection. A review of systems should focus on any neurologic signs or sensorium alterations, rashes or skin lesions, and focal pain or tenderness and should also include a general review of respiratory, gastrointestinal, or genitourinary symptoms.

PHYSICAL EXAMINATION

A complete physical examination should be performed, with special attention to several areas that are sometimes given short shrift in routine examinations. Assessment of the patient's general appearance and vital signs, skin and soft tissue examination, and the neurologic evaluation are of particular importance.

The patient may appear either anxious and agitated or lethargic and apathetic. Fever is usually present, although elderly patients and compromised hosts (e.g., patients who are uremic or cirrhotic and those who are taking glucocorticoids or nonsteroidal anti-inflammatory drugs) may be afebrile despite serious underlying infection. Critically ill patients may be hypothermic, with a high risk of organ failure and mortality. Measurement of blood pressure, heart rate, and respiratory rate helps determine the degree of hemodynamic and metabolic compromise. The patient's airway must be evaluated to rule out the risk of obstruction from an invasive oropharyngeal infection.

The etiologic diagnosis may become evident in the context of a thorough skin examination (**Chap. 16**). Petechial rashes are typically seen with meningococcemia or Rocky Mountain spotted fever (RMSF; see **Fig. A1-16**); erythroderma is associated with toxic shock syndrome (TSS) and drug fever. The soft tissue and muscle examination is critical. Areas of erythema or duskiness, edema, and tenderness may indicate underlying necrotizing fasciitis, myositis, or myonecrosis. The neurologic examination must include a careful assessment of mental status for signs of early encephalopathy. Evidence of nuchal rigidity or focal neurologic findings should be sought.

DIAGNOSTIC WORKUP

After a quick clinical assessment, diagnostic material should be obtained rapidly and antibiotic and supportive treatment begun. Blood (for cultures; baseline complete blood count with differential; measurement of serum electrolytes, blood urea nitrogen, serum creatinine, and serum glucose; and liver function tests) can be obtained at the time an IV line is placed and before antibiotics are administered. The blood lactate concentration also should be measured. Three sets of blood cultures should be performed for patients with possible acute endocarditis. Blood smears from patients at risk for severe parasitic disease, such as malaria or babesiosis (**Chaps. 219, 220, and A6**), must be examined for the diagnosis and quantitation of parasitemia. Blood smears may also be diagnostic in ehrlichiosis and anaplasmosis.

Patients with possible meningitis should have cerebrospinal fluid (CSF) drawn before the initiation of antibiotic therapy. Focal findings, depressed mental status, or papilledema should be evaluated by brain imaging prior to lumbar puncture, which, in this setting, could initiate herniation. *Antibiotics should be administered before imaging but after blood for cultures has been drawn.* If CSF cultures are negative, blood

cultures will provide the diagnosis in 50–70% of cases. Molecular diagnostic techniques (e.g., broad-range 16S rRNA gene polymerase chain reaction testing for bacterial meningitis pathogens) are of increasing importance in the rapid diagnosis of life-threatening infections.

Focal abscesses necessitate immediate CT or MRI as part of an evaluation for surgical intervention. Other diagnostic procedures, such as wound cultures, should not delay the initiation of treatment for more than minutes. Once emergent evaluation, diagnostic procedures, and (if appropriate) surgical consultation (see below) have been completed, other laboratory tests can be conducted. Appropriate radiography, computed axial tomography, MRI, urinalysis, measurement of the erythrocyte sedimentation rate and/or C-reactive protein level, procalcitonin monitoring, and transthoracic or transesophageal echocardiography all may prove important.

TREATMENT

TREATMENT

The Acutely Ill Patient

In the acutely ill patient, empirical antibiotic therapy is critical and should be administered without undue delay in addition to fluid resuscitation and vasopressor support as needed. Increased prevalence of antibiotic resistance in community-acquired bacteria must be considered when antibiotics are selected.

Table 117-1 lists first-line empirical regimens for infections considered in this chapter. In addition to the rapid initiation of antibiotic therapy, several of these infections require urgent surgical attention. Neurosurgical evaluation for subdural empyema, otolaryngologic surgery for possible mucormycosis, and cardiothoracic surgery for critically ill patients with acute endocarditis are as important as antibiotic therapy. For infections such as necrotizing fasciitis and clostridial myonecrosis, rapid surgical intervention supersedes other diagnostic or therapeutic maneuvers.

Adjunctive treatments may reduce morbidity and mortality rates and include **dexamethasone** for bacterial meningitis or IV immunoglobulin for TSS and necrotizing fasciitis caused by group A *Streptococcus*.

Adjunctive therapies should usually be initiated within the first hours of treatment; however, **dexamethasone** for bacterial meningitis must be given before or at the time of the first dose of antibiotic. Glucocorticoids can also be harmful, sometimes resulting in worse outcomes—e.g., when given in the setting of cerebral malaria or viral hepatitis.

TABLE 117-1

Empirical Treatment for Common Infectious Disease Emergencies^a

CLINICAL SYNDROME	POSSIBLE ETIOLOGIES	TREATMENT	COMMENTS	SEE CHAP(S).
Sepsis without a Clear Focus				
Septic shock	<i>Pseudomonas</i> spp., gram-negative enteric bacilli, <i>Staphylococcus</i> spp., <i>Streptococcus</i> spp.	Vancomycin (15 mg/kg q12h) ^b <i>plus</i> gentamicin (5 mg/kg per day) <i>plus either</i> Piperacillin/tazobactam (3.375–4.5 g q6h) <i>or</i> cefepime (2 g q8h) ^c	Empirical therapy should be tailored to local resistance patterns. Adjust treatment when culture data become available.	142, 143, 156, 159, 297
Overwhelming post-splenectomy sepsis	<i>Streptococcus pneumoniae</i> , <i>Haemophilus influenzae</i> , <i>Neisseria meningitidis</i>	Ceftriaxone (2 g q12h) <i>plus</i> vancomycin (15 mg/kg q12h) ^b	If a β -lactam-sensitive strain is identified, vancomycin can be discontinued.	297
Babesiosis	<i>Babesia microti</i> (U.S.), <i>B. divergens</i> (Europe)	Clindamycin (600 mg q8h) <i>plus</i> quinine (650 mg q8h)	Atovaquone and azithromycin can be used in less severe disease and are associated with fewer side effects. Treatment with doxycycline (100 mg bid) for potential co-infection with <i>Borrelia burgdorferi</i> or <i>Anaplasma</i> spp. may be prudent.	217, 220
Sepsis with Skin Findings				

CLINICAL SYNDROME	POSSIBLE ETIOLOGIES	TREATMENT	COMMENTS	SEE CHAP(S).
Meningococccemia	<i>N. meningitidis</i>	Penicillin (4 mU q4h) <i>or</i> ceftriaxone (2 g q12h)	Ceftriaxone eradicates nasopharyngeal carriage of the organism. Close contacts require chemoprophylaxis with rifampin (600 mg q12h for 2 days) or ciprofloxacin (a single dose, 500 mg).	150
Rocky Mountain spotted fever (RMSF)	<i>Rickettsia rickettsii</i>	Doxycycline (100 mg bid)	If both meningococccemia and RMSF are being considered, use ceftriaxone (2 g q12h) <i>plus</i> doxycycline (100 mg bid). If RMSF is diagnosed, doxycycline is the proven superior agent.	182
Purpura fulminans	<i>S. pneumoniae</i> , <i>H. influenzae</i> , <i>N. meningitidis</i>	Ceftriaxone (2 g q12h) <i>plus</i> vancomycin (15 mg/kg q12h) ^b	If a β -lactam-sensitive strain is identified, vancomycin can be discontinued.	141, 150, 152, 297
Erythroderma: toxic shock syndrome	Group A <i>Streptococcus</i> , <i>Staphylococcus aureus</i>	Vancomycin (15 mg/kg q12h) ^b <i>plus</i> clindamycin (600 mg q8h)	If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 2 mU q4h; or oxacillin, 2 g IV q4h). The site of toxigenic bacteria should be debrided; IV immunoglobulin can be used in severe cases. ^d	142, 143

CLINICAL SYNDROME	POSSIBLE ETIOLOGIES	TREATMENT	COMMENTS	SEE CHAP(S).
Sepsis with Soft Tissue Findings				
Necrotizing fasciitis	Group A <i>Streptococcus</i> , mixed aerobic/anaerobic flora, CA-MRSA ^e	Vancomycin (15 mg/kg q12h) ^b <i>plus</i> clindamycin (600 mg q8h) <i>plus</i> gentamicin (5 mg/kg per day)	Urgent surgical evaluation is critical. Adjust treatment when culture data become available.	124, 142, 143
Clostridial myonecrosis	<i>Clostridium perfringens</i>	Penicillin (2 mU q4h) <i>plus</i> clindamycin (600 mg q8h)	Urgent surgical evaluation is critical.	149
Neurologic Infections				
Bacterial meningitis	<i>S. pneumoniae</i> , <i>N. meningitidis</i>	Ceftriaxone (2 g q12h) <i>plus</i> vancomycin (15 mg/kg q12h) ^b	If a β -lactam-sensitive strain is identified, vancomycin can be discontinued. If the patient is >50 years old or has comorbid disease, add ampicillin (2 g q4h) for <i>Listeria</i> coverage. Dexamethasone (10 mg q6h for 4 days) improves outcome in adults with meningitis (especially pneumococcal).	133
Brain abscess, suppurative intracranial infections	<i>Streptococcus</i> spp., <i>Staphylococcus</i> spp., anaerobes, gram-negative bacilli	Vancomycin (15 mg/kg q12h) ^b <i>plus</i> metronidazole (500 mg q8h) <i>plus</i> ceftriaxone (2 g q12h)	Urgent surgical evaluation is critical. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 4 mU q4h; or oxacillin, 2 g q4h).	133

CLINICAL SYNDROME	POSSIBLE ETIOLOGIES	TREATMENT	COMMENTS	SEE CHAP(S).
Cerebral malaria	<i>Plasmodium falciparum</i>	Artesunate (2.4 mg/kg IV at 0, 12, and 24 h; then once daily) ^f <i>or</i> quinine (IV loading dose of 20 mg salt/kg; then 10 mg/kg q8h)	Do not use glucocorticoids. Use IV quinidine if IV quinine is not available. During IV quinidine treatment, blood pressure and cardiac function should be monitored continuously and blood glucose periodically.	217, 219
Spinal epidural abscess	<i>Staphylococcus</i> spp., gram-negative bacilli	Vancomycin (15 mg/kg q12h) ^b <i>plus either</i> Piperacillin/tazobactam (3.375–4.5 g q6h) <i>or</i> cefepime (2 g q8h) ^c	Surgical evaluation is essential. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 4 mU q4h; <i>or</i> oxacillin, 2 g q4h).	434
Focal Infections				
Acute bacterial endocarditis	<i>S. aureus</i> , β -hemolytic streptococci, HACEK group, ^g <i>Neisseria</i> spp., <i>S. pneumoniae</i>	Ceftriaxone (2 g q12h) <i>plus</i> vancomycin (15 mg/kg q12h) ^b	Adjust treatment when culture data become available. Surgical evaluation is essential.	123

a

These empirical regimens include coverage for gram-positive pathogens that are resistant to β -lactam antibiotics. Local resistance patterns should be considered and may alter the need for empirical vancomycin.

b

A vancomycin loading dose of 20–25 mg/kg can be considered in critically ill patients.

c

β -Lactam antibiotics may exhibit unpredictable pharmacodynamics in sepsis. Higher dosing or prolonged or continuous infusions can be considered.

d

The optimal dose of IV immunoglobulin has not been determined, but the median dose in observational studies is 2 g/kg (total dose administered for 1–5 days).

e

Community-acquired methicillin-resistant *S. aureus*.

f

In the United States, artesunate must be obtained through the Centers for Disease Control and Prevention. For patients diagnosed with severe malaria, full doses of parenteral antimalarial treatment should be started with whichever recommended antimalarial agent is first available.

g

Haemophilus spp., *Aggregatibacter* spp., *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*.

SPECIFIC PRESENTATIONS

The infections considered below according to common clinical presentation can have rapidly catastrophic outcomes, and their immediate recognition and treatment can be life-saving. Recommended empirical therapeutic regimens are presented in [Table 117-1](#).

SEPSIS WITHOUT AN OBVIOUS FOCUS OF PRIMARY INFECTION

Patients initially have a brief prodrome of nonspecific symptoms and signs that progresses quickly to hemodynamic instability with hypotension, tachycardia, tachypnea, respiratory distress, and altered mental status. Disseminated intravascular coagulation (DIC) with clinical evidence of a hemorrhagic diathesis is a poor prognostic sign.

Septic Shock

Patients with bacteremia leading to septic shock may have a primary site of infection (e.g., pneumonia, pyelonephritis, or cholangitis) that is not evident initially (See also [Chap. 297](#)). Elderly patients with comorbid conditions, hosts compromised by malignancy and neutropenia, and patients who have recently undergone a surgical procedure or hospitalization are at increased risk for an adverse outcome. Gram-negative bacteremia with organisms such as *Pseudomonas aeruginosa* or *Escherichia coli* and gram-positive infection with organisms such as *Staphylococcus aureus* (including methicillin-resistant *S. aureus* [MRSA]) or group A streptococci can present as intractable hypotension and multiorgan failure. Treatment can usually be initiated empirically on the basis of the presentation, host factors ([Chap. 297](#)), and local patterns of bacterial resistance. Outcomes are worse when antimicrobial treatment is delayed or when the responsible pathogen ultimately proves not to be susceptible to the initial regimen. Active empirical antimicrobial coverage administered before admission to the intensive care unit is strongly associated with improved

survival. Broad-spectrum antimicrobial agents are therefore recommended and should be instituted rapidly, preferably within the first hours after presentation. Risk factors for fungal infection should be assessed, as the incidence of fungal septic shock is increasing. Biomarkers such as C-reactive protein and procalcitonin have not proved reliable diagnostically but, when measured over time, can facilitate appropriate de-escalation of therapy and predict outcome. Glucocorticoids are often considered for patients with severe sepsis who do not respond to fluid resuscitation and vasopressor therapy. However, conclusive evidence for the efficacy of glucocorticoids in this setting is lacking.

Overwhelming Infection in Asplenic Patients

Patients without splenic function are at risk for overwhelming bacterial sepsis (See also [Chap. 297](#)). Asplenic adult patients succumb to sepsis at 58 times the rate of the general population. Most infections are thought to occur within the first 1 or 2 years, but the increased risk persists throughout life. The median interval between splenectomy and sepsis is 5.75 years, with a range of 1–19 years. In asplenia, encapsulated bacteria cause the majority of infections. Adults, who are more likely to have antibody to these organisms, are at lower risk than children. *Streptococcus pneumoniae* is the most common isolate, causing 40–70% of cases. The risk of infection with *Haemophilus influenzae* or *Neisseria meningitidis* is also greater in patients without splenic function, but reported cases are declining. Severe clinical manifestations of infections due to *E. coli*, *S. aureus*, group B streptococci, *P. aeruginosa*, *Bordetella holmesii*, and *Capnocytophaga*, *Babesia*, and *Plasmodium* species have been described.

Babesiosis

A history of recent travel to endemic areas raises the possibility of infection with *Babesia* (See also [Chap. 220](#)). Between 1 and 4 weeks after a tick bite, the patient experiences chills, fatigue, anorexia, myalgia, arthralgia, shortness of breath, nausea, and headache; ecchymosis and/or petechiae are occasionally seen. The tick that most commonly transmits *Babesia*, *Ixodes scapularis*, also transmits *Borrelia burgdorferi* (the agent of Lyme disease) and *Anaplasma*; co-infection can occur, resulting in more severe disease. Infection with the European species *Babesia divergens* is more frequently fulminant than that due to the U.S. species *Babesia microti*. *B. divergens* causes a febrile syndrome with hemolysis, jaundice, hemoglobinemia, and renal failure and is associated with a mortality rate of >40%. Severe babesiosis is especially common in asplenic hosts but does occur in hosts with normal splenic function, particularly those >60 years of age and those with underlying immunosuppressive conditions such as HIV infection or malignancy. Complications include renal failure, acute respiratory failure, and DIC.

Other Sepsis Syndromes

Tularemia ([Chap. 165](#)) is seen throughout the United States, but most cases recorded in 2015 occurred in South Dakota, Nebraska, Colorado, and Wyoming. This disease is associated with wild rabbit, tick, and tabanid fly contact. It can be transmitted by arthropod bite, handling of infected animal carcasses, consumption of contaminated food and water, or inhalation. The typhoidal form can be associated with gram-negative septic shock and a mortality rate of >30%, especially in patients with underlying comorbid or

immunosuppressive conditions. Plague occurs infrequently in the United States (**Chap. 166**), primarily after contact with ground squirrels, prairie dogs, or chipmunks, but is endemic in other parts of the world, with >90% of all cases occurring in Africa. The septic form is particularly rare and is associated with shock, multiorgan failure, and a 30% mortality rate. These infections should be considered in the appropriate epidemiologic setting. The Centers for Disease Control and Prevention lists *Francisella tularensis* and *Yersinia pestis* (the agents of tularemia and plague, respectively) along with *Bacillus anthracis* (the agent of anthrax) as important organisms that might be used for bioterrorism (**Chap. S2**).

SEPSIS WITH SKIN MANIFESTATIONS

Maculopapular rashes may reflect early meningococcal or rickettsial disease but are usually associated with nonemergent infections (See also **Chap. 16**). Exanthems are usually viral. Primary HIV infection commonly presents with a rash that is typically maculopapular and involves the upper part of the body but can spread to the palms and soles. The patient is usually febrile and can have lymphadenopathy, severe headache, dysphagia, diarrhea, myalgias, and arthralgias. Recognition of this syndrome provides an opportunity to prevent transmission and to institute treatment and monitoring early on.

Petechial rashes caused by viruses are seldom associated with hypotension or a toxic appearance, although there can be exceptions (e.g., severe measles or arboviral infection). Petechial rashes limited to the distribution of the superior vena cava are rarely associated with severe disease. In other settings, petechial rashes require more urgent attention.

Meningococcemia



Almost three-quarters of patients with *N. meningitidis* bacteremia have a rash (See also **Chap. 150**). Meningococcemia most often affects young children (i.e., those 6 months to 5 years old). In sub-Saharan Africa, the high prevalence of serogroup A meningococcal disease has been a threat to public health for more than a century. Thousands of deaths occur annually in this area, which is known as the “meningitis belt,” and large epidemic waves occur approximately every 8–12 years. Serogroups W135 and X are also important emerging pathogens in Africa. In the United States, sporadic cases and outbreaks occur in day-care centers, schools (grade school through college, particularly among college freshmen living in residential halls), and army barracks. Household contacts of index cases are at 400–800 times greater risk of disease than the general population. Patients may have fever, headache, nausea, vomiting, myalgias, changes in mental status, and meningismus. However, the rapidly progressive form of disease is not usually associated with meningitis. The rash is initially pink, blanching, and maculopapular, appearing on the trunk and extremities, but then becomes hemorrhagic, forming petechiae. Petechiae are first seen at the ankles, wrists, axillae, mucosal surfaces, and palpebral and bulbar conjunctiva, with subsequent spread on the lower extremities and to the trunk. A cluster of petechiae may be seen at pressure points—e.g., where a blood pressure cuff has been inflated. In rapidly progressive meningococcemia (10–20% of cases), the petechial rash quickly becomes purpuric (see **Fig. A1-41**), and patients develop DIC, multiorgan failure, and shock; 50–60% of these patients die, and survivors often require extensive debridement or amputation of gangrenous extremities.

Hypotension with petechiae for <12 h is associated with significant mortality. Cyanosis, coma, oliguria, metabolic acidosis, and elevated partial thromboplastin time also are associated with a fatal outcome. Antibiotics given in the office by the primary care provider before hospital evaluation and admission may improve prognosis; this observation suggests that early initiation of treatment may be life-saving. Meningococcal conjugate vaccines are protective against serogroups A, C, Y and W135 and are recommended for children 11–18 years of age and for other high-risk patients. Vaccines active against serogroup B are available and are recommended for high-risk individuals >10 years of age.

Rocky Mountain Spotted Fever and Other Rickettsial Diseases

RMSF is a tickborne disease caused by *Rickettsia rickettsii* that occurs throughout North and South America (See also [Chap. 182](#)). Other rickettsiae (e.g., *R. parkeri*, *R. akari*) can also cause spotted fever. Up to 40% of patients do not report a history of a tick bite, but a history of travel or outdoor activity (e.g., camping in tick-infested areas) can often be ascertained. For the first 3 days, headache, fever, malaise, myalgias, nausea, vomiting, and anorexia are documented. By day 3, half of patients have skin findings. Blanching macules develop initially on the wrists and ankles and then spread over the legs and trunk. The lesions become hemorrhagic and are frequently petechial. The rash spreads to palms and soles later in the course. The centripetal spread is a classic feature of RMSF but occurs in a minority of patients. Moreover, 10–15% of patients with RMSF never develop a rash. The patient can be hypotensive and develop noncardiogenic pulmonary edema, confusion, lethargy, and encephalitis progressing to coma. The CSF contains 10–100 cells/ μ L, usually with a predominance of mononuclear cells. The CSF glucose level is often normal; the protein concentration may be slightly elevated. Renal and hepatic injury as well as bleeding secondary to vascular damage are noted. For untreated infections, mortality rates are 20–30%. Delayed recognition and treatment are associated with a greater risk of death; Native Americans, children 5–9 years of age, adults >70 years old, and persons with underlying immunosuppression are at a 3- to 5-fold increased risk of death.



Other rickettsial diseases cause significant morbidity and mortality worldwide. *Mediterranean spotted fever* caused by *Rickettsia conorii* is found in Africa, southwestern and south-central Asia, and southern Europe. Patients have fever, flu-like symptoms, and an inoculation eschar at the site of the tick bite. A maculopapular rash develops within 1–7 days, involving the palms and soles but sparing the face. Elderly patients or those with diabetes, alcoholism, uremia, or congestive heart failure are at risk for severe disease characterized by neurologic involvement, respiratory distress, and gangrene of the digits or purpura fulminans. Mortality rates associated with this severe form of disease approach 50%. *Epidemic typhus*, caused by *Rickettsia prowazekii*, is transmitted in louse-infested environments and emerges in conditions of extreme poverty, war, and natural disaster. Patients experience a sudden onset of high fevers, severe headache, cough, myalgias, and abdominal pain. A maculopapular rash develops (primarily on the trunk) in more than half of patients and can progress to petechiae and purpura. Serious signs include delirium, coma, seizures, noncardiogenic pulmonary edema, skin necrosis, and peripheral gangrene. Mortality rates approached 60% in the preantibiotic era and continue to exceed 10–15% in contemporary outbreaks. *Scrub typhus*, caused by *Orientia tsutsugamushi* (a separate genus in the family Rickettsiaceae), is transmitted by larval mites or chiggers and is one of the most common infections in southeastern Asia and the western

Pacific. The organism is found in areas of heavy scrub vegetation (e.g., along riverbanks). Patients may have an inoculation eschar and may develop a maculopapular rash, lymphadenopathy, and dyspnea. Severe cases progress to pneumonia, meningoencephalitis, myocarditis, DIC, and renal failure. Mortality rates range from 1% to 70% and vary by location, increasing age, myocarditis, delirium, pneumonitis, or signs of hemorrhage.

If recognized in a timely fashion, rickettsial disease is very responsive to treatment. Doxycycline (100 mg twice daily for 3–14 days) is the treatment of choice for both adults and children. The newer macrolides may be a suitable alternative, but mortality rates are higher when tetracycline-based treatment is not given.

Purpura Fulminans

Purpura fulminans is the cutaneous manifestation of DIC and presents as large ecchymotic areas and hemorrhagic bullae (See also [Chaps. 150](#) and [297](#)). Progression of petechiae to purpura, ecchymoses, and gangrene is associated with congestive heart failure, septic shock, acute renal failure, acidosis, hypoxia, hypotension, and death. Purpura fulminans has been associated primarily with *N. meningitidis* but, in splenectomized patients, may be associated with *S. pneumoniae*, *H. influenzae*, and *S. aureus*.

Ecthyma Gangrenosum

Septic shock caused by *P. aeruginosa* or *Aeromonas hydrophila* can be associated with ecthyma gangrenosum (see [Figs. 159-1](#) and [A1-34](#)): hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration. These gram-negative bacteremias are most common among patients with neutropenia, extensive burns, and hypogammaglobulinemia.

Other Infections Associated with Rash

Vibrio vulnificus and other noncholera *Vibrio* bacteremic infections ([Chap. 163](#)) can cause focal skin lesions and overwhelming sepsis in hosts with chronic liver disease, heavy alcohol consumption, iron storage disorders, diabetes, renal insufficiency, hematologic disease, or malignancy or other immunocompromising conditions. After ingestion of contaminated raw shellfish (typically oysters from the Gulf Coast in U.S. cases), there is a sudden onset of malaise, chills, fever, and hypotension. The patient develops bullous or hemorrhagic skin lesions, usually on the lower extremities, and 75% of patients have leg pain. The mortality rate can be as high as 50–60%, particularly when the patient presents with hypotension. Outcomes are improved when patients are treated with fluoroquinolones with or without cephalosporins or with tetracycline-containing regimens. Other infections, caused by agents such as *Aeromonas*, *Klebsiella*, and *E. coli*, can cause hemorrhagic bullae and death due to overwhelming sepsis in cirrhotic patients.

Capnocytophaga canimorsus can cause septic shock in asplenic patients. Infection typically follows a dog bite. Patients present with fever, chills, myalgia, vomiting, diarrhea, dyspnea, confusion, and headache. Findings can include an exanthem or erythema multiforme (see [Figs. 52-9](#) and [A1-24](#)), cyanotic mottling or peripheral cyanosis, petechiae, and ecchymosis. About 30% of patients with this fulminant form die of overwhelming sepsis and DIC, and survivors may require amputation because of gangrene.

Erythroderma

TSS (**Chaps. 142 and 143**) is usually associated with erythroderma. The patient presents with fever, malaise, myalgias, nausea, vomiting, diarrhea, and confusion. There is a sunburn-type rash that may be subtle and patchy but is usually diffuse and is found on the face, trunk, and extremities. Erythroderma, which desquamates after 1–2 weeks, is more common in *Staphylococcus*-associated than in *Streptococcus*-associated TSS. Hypotension develops rapidly—often within hours—after the onset of symptoms. Multiorgan failure occurs. Early renal failure may precede hypotension and distinguishes this syndrome from other septic shock syndromes. There may be no indication of a primary focal infection, although possible cutaneous or mucosal portals of entry for the organism can be ascertained when a careful history is taken. Colonization rather than overt infection of the vagina or a postoperative wound, for example, is typical with staphylococcal TSS, and the mucosal areas appear hyperemic but not infected. Streptococcal TSS is more often associated with skin or soft tissue infection (including necrotizing fasciitis), and patients are more likely to be bacteremic. TSS caused by *Clostridium sordellii* is associated with childbirth or with skin injection of black-tar heroin. The diagnosis of TSS is defined by the clinical criteria of fever, rash, hypotension, and multiorgan involvement. (Of note, fever is typically absent when TSS is caused by *C. sordellii*.) The mortality rate is 5% for menstruation-associated TSS, 10–15% for nonmenstrual TSS, 30–70% for streptococcal TSS, and up to 90% for obstetric *C. sordellii* TSS. **Clindamycin** improves outcomes when included in the treatment regimen. Some studies have shown that use of IV immunoglobulin is associated with improved survival as well.

Viral Hemorrhagic Fevers



Viral hemorrhagic fevers (**Chaps. 204 and 205**) are zoonotic illnesses caused by viruses that reside in either animal reservoirs or arthropod vectors. These diseases occur worldwide and are restricted to areas where the host species live. They are caused by four major groups of viruses: Arenaviridae (e.g., Lassa fever in Africa), Bunyaviridae (e.g., Rift Valley fever in Africa; hantavirus hemorrhagic fever with renal syndrome in Asia; and Crimean-Congo hemorrhagic fever, which has an extensive geographic distribution), Filoviridae (e.g., Ebola and Marburg virus infections in Africa), and Flaviviridae (e.g., yellow fever in Africa and South America and dengue in Asia, Africa, and the Americas). Lassa fever and Ebola and Marburg virus infections are also transmitted from person to person. The vectors for most viral fevers are found in rural areas; dengue and yellow fever are important exceptions. After a prodrome of fever, myalgias, and malaise, patients develop evidence of vascular damage, petechiae, and local hemorrhage. Shock, multifocal hemorrhaging, and neurologic signs (e.g., seizures or coma) predict a poor prognosis. Dengue (**Chap. 204**) is the most common arboviral disease worldwide. More than half a million cases of dengue hemorrhagic fever occur each year, with at least 12,000 deaths. Patients have a triad of symptoms: hemorrhagic manifestations, evidence of plasma leakage, and platelet counts of $<100,000/\mu\text{L}$. Mortality rates are 10–20%. If dengue shock syndrome develops, mortality rates can reach 40%. Ebola infection has been associated with outbreaks with high mortality rates. The 2014 outbreak in West Africa had a mortality rate of $>50\%$. Symptoms can appear 2–21 days after exposure, but most patients become ill within 9 days. The patient first presents with fatigue, fever, headache, and muscle pains, and the illness can progress to multiorgan failure and hemorrhaging.

Careful volume-replacement therapy to maintain blood pressure and intravascular volume is key to survival in these infections. [Ribavirin](#) also may be useful against Arenaviridae and Bunyaviridae.

Other viral illnesses with rash, such as measles, can be associated with significant mortality rates. Steroids may sometimes be useful in severe disease in malnourished populations, especially if neurologic complications are present.

SEPSIS WITH A SOFT TISSUE/MUSCLE PRIMARY FOCUS

See also [Chap. 124](#).

Necrotizing Fasciitis

This infection is characterized by extensive necrosis of the subcutaneous tissue and fascia. It may arise at a site of minimal trauma or surgical incision and may also be associated with recent varicella, childbirth, or muscle strain. The most common causes of necrotizing fasciitis are group A streptococci alone ([Chap. 143](#)) and a mixed facultative and anaerobic flora ([Chap. 124](#)); the incidence of group A streptococcal necrotizing fasciitis has been increasing for the past quarter-century. Diabetes mellitus, IV drug use, chronic liver or renal disease, and malignancy are associated risk factors. Physical findings are initially minimal compared with the severity of pain and the degree of fever. The examination is often unremarkable except for soft tissue edema and erythema. The infected area is red, hot, shiny, swollen, and exquisitely tender. In untreated infection, the overlying skin develops blue-gray patches after 36 h, and cutaneous bullae and necrosis develop after 3–5 days. Necrotizing fasciitis due to a mixed flora, but not that due to group A streptococci, can be associated with gas production. Without treatment, pain decreases because of thrombosis of the small blood vessels and destruction of the peripheral nerves—an ominous sign. The mortality rate is 15–34% overall, >70% in association with TSS, and nearly 100% without surgical intervention. Necrotizing fasciitis may also be due to *Clostridium perfringens* ([Chap. 149](#)); in this condition, the patient is extremely toxic and the mortality rate is high. Within 48 h, rapid tissue invasion and systemic toxicity associated with hemolysis and death ensue. The distinction between this entity and clostridial myonecrosis is made by muscle biopsy. Necrotizing fasciitis caused by community-acquired MRSA also has been reported.

Clostridial Myonecrosis

Myonecrosis is often associated with trauma or surgery but can develop spontaneously (See also [Chap. 149](#)). The incubation period is usually 12–24 h long, and massive necrotizing gangrene develops within hours of onset. Systemic toxicity, shock, and death can occur within 12 h. The patient's pain and toxic appearance are out of proportion to physical findings. On examination, the patient is febrile, apathetic, tachycardic, and tachypneic and may express a feeling of impending doom. Hypotension and renal failure develop later, and hyperalertness is evident preterminally. The skin over the affected area is bronze-brown, mottled, and edematous. Bullous lesions with serosanguineous drainage and a mousy or sweet odor can develop. Crepitus can occur secondary to gas production in muscle tissue. The mortality rate is >65% for spontaneous myonecrosis, which is often associated with *Clostridium septicum* or *C. tertium* and underlying malignancy.

The mortality rates associated with trunk and limb infection are 63% and 12%, respectively, and any delay in surgical treatment increases the risk of death.

NEUROLOGIC INFECTIONS WITH OR WITHOUT SEPTIC SHOCK

Bacterial Meningitis

Bacterial meningitis is one of the most common infectious disease emergencies involving the central nervous system (See also [Chap. 133](#)). Although hosts with cell-mediated immune deficiency (including transplant recipients, diabetic patients, elderly patients, and cancer patients receiving certain chemotherapeutic agents) are at particular risk for *Listeria monocytogenes* meningitis, most cases in adults are due to *S. pneumoniae* (30–60%) and *N. meningitidis* (10–35%). The classic presentation of fever, meningismus, and altered mental status is seen in only one-half to two-thirds of patients. The elderly can present without fever or meningeal signs. Cerebral dysfunction is evidenced by confusion, delirium, and lethargy that can progress to coma. In some cases, the presentation is fulminant, with sepsis and brain edema; papilledema at presentation is unusual and suggests another diagnosis (e.g., an intracranial lesion). Focal signs, including cranial nerve palsies (IV, VI, VII), can be seen in 10–20% of cases; 50–70% of patients have bacteremia. A poor outcome is associated with coma, seizures, hypotension, a pneumococcal etiology, respiratory distress, a CSF glucose level of <0.6 mmol/L (<10 mg/dL), a CSF protein level of >2.5 g/L, a peripheral white blood cell count of $<5000/\mu\text{L}$, and a serum sodium level of <135 mmol/L. Rapid initiation of treatment is essential; the odds of an unfavorable outcome may increase by 30% for each hour that treatment is delayed.

[Dexamethasone](#) is an adjunctive treatment for meningitis in adults, especially for infections caused by *S. pneumoniae*. It must be given before or with the first dose of antibiotics; otherwise, it is unlikely to improve outcomes.

Suppurative Intracranial Infections

In suppurative intracranial infections, rare intracranial lesions present along with sepsis and hemodynamic instability (See also [Chap. 135](#)). Rapid recognition of the toxic patient with central neurologic signs is crucial to improvement of the dismal prognosis of these entities. Patients with diabetes or hematologic disease may be at increased risk for these infections. *Subdural empyema* arises from the paranasal sinus in 60–70% of cases. Microaerophilic streptococci and staphylococci are the predominant etiologic organisms. The patient is toxic, with fever, headache, and nuchal rigidity. Of all patients, 75% have focal signs and 6–20% die. Despite improved survival rates, 15–44% of patients are left with permanent neurologic deficits. *Septic cavernous sinus thrombosis* follows a facial or sphenoid sinus infection; 70% of cases are due to staphylococci (including MRSA), and the remainder are due primarily to aerobic or anaerobic streptococci. Fungi have been common in some series. A unilateral or retro-orbital headache progresses to a toxic appearance and fever within days. Three-quarters of patients have unilateral periorbital edema that becomes bilateral and then progresses to ptosis, proptosis, ophthalmoplegia, and papilledema. The mortality rate is as high as 30%. *Septic thrombosis of the superior sagittal sinus* spreads from the ethmoid or maxillary sinuses and is caused by *S. pneumoniae*, other streptococci, and staphylococci. The fulminant course is characterized by headache, nausea, vomiting, rapid progression to confusion and coma, nuchal

rigidity, and brainstem signs. If the sinus is totally thrombosed, the mortality rate exceeds 80%. Broad-spectrum antibiotics and early surgical intervention at the primary site of infection may improve outcomes. Anticoagulation or steroids are of uncertain benefit.

Brain Abscess

Brain abscess often occurs without systemic signs (See also [Chap. 135](#)). Almost half of patients are afebrile, and presentations are more consistent with a space-occupying lesion in the brain; 70% of patients have headache and/or altered mental status, 50% have focal neurologic signs, and 25% have papilledema. Abscesses can present as single or multiple lesions resulting from contiguous foci or hematogenous infection, such as endocarditis, or after surgery or trauma. The infection progresses over several days from cerebritis to an abscess with a mature capsule. More than half of infections are polymicrobial, with an etiology consisting of aerobic bacteria (primarily streptococcal species) and anaerobes. Abscesses arising hematogenously are especially apt to rupture into the ventricular space, causing a sudden and severe deterioration in clinical status and a high mortality rate. Otherwise, mortality is low (<20%) but morbidity is high (30–55%). Patients presenting with stroke and a parameningeal infectious focus, such as sinusitis or otitis, may have a brain abscess, and physicians must maintain a high level of suspicion. Prognosis worsens in patients with a fulminant course, delayed diagnosis, abscess rupture into the ventricles, multiple abscesses, or abnormal neurologic status at presentation. In one study, mortality at 1 year was 19%.

Cerebral Malaria

This entity should be urgently considered if patients who have recently traveled to areas endemic for malaria present with a febrile illness and lethargy or other neurologic signs (See also [Chap. 219](#)). Fulminant malaria is caused by *Plasmodium falciparum* and is associated with temperatures of >40°C (>104°F), hypotension, jaundice, acute respiratory distress syndrome, and bleeding. By definition, any patient with a change in mental status or repeated seizure in the setting of fulminant malaria has cerebral malaria. In adults, this nonspecific febrile illness progresses to coma over several days; occasionally, coma occurs within hours and death within 24 h. Nuchal rigidity and photophobia are rare. On physical examination, symmetric encephalopathy is typical, and upper motor neuron dysfunction with decorticate and decerebrate posturing can be seen in advanced disease. Unrecognized infection results in a 20–30% mortality rate.

Intracranial and Spinal Epidural Abscesses

Spinal and intracranial epidural abscesses (SEAs and ICEAs) can result in permanent neurologic deficits, sepsis, and death (See also [Chap. 434](#)). At-risk patients include those with diabetes mellitus; IV drug use; chronic alcohol abuse; recent spinal trauma, surgery, or epidural anesthesia; and other comorbid conditions, such as HIV infection. Fungal epidural abscess and meningitis can follow epidural or paraspinal glucocorticoid injections. In the United States and Canada, where early treatment of otitis and sinusitis is typical, ICEA is rare but the number of cases of SEA is on the rise. In Africa and areas with limited access to health care, SEAs and ICEAs cause significant morbidity and mortality. ICEAs typically present as fever, mental status changes, and neck pain, while SEAs often present as fever, localized spinal tenderness, and

back pain. ICEAs are typically polymicrobial, whereas SEAs are most often due to hematogenous seeding, with staphylococci the most common etiologic agent. Early diagnosis and treatment, which may include surgical drainage, minimize rates of mortality and permanent neurologic sequelae. Outcomes are worse for SEA due to MRSA, for infection at a higher vertebral-body level, for impaired neurologic status on presentation, and for dorsal rather than ventral location of the abscess. Elderly patients and persons with renal failure, malignancy, and other comorbidities also have less favorable outcomes.

OTHER FOCAL SYNDROMES WITH A FULMINANT COURSE

Infection at virtually any primary focus (e.g., osteomyelitis, pneumonia, pyelonephritis, or cholangitis) can result in bacteremia and sepsis. Lemierre's syndrome—jugular septic thrombophlebitis caused by *Fusobacterium necrophorum*—is associated with metastatic infectious emboli (primarily to the lung but sometimes to the liver or other organs) and sepsis, with mortality rates of >15%. TSS has been associated with focal infections such as septic arthritis, peritonitis, sinusitis, and wound infection. Rapid clinical deterioration and death can be associated with destruction of the primary site of infection, as is seen in endocarditis and in infections of the oropharynx (e.g., Ludwig's angina or epiglottitis, in which edema suddenly compromises the airway).

Rhinocerebral Mucormycosis

Individuals with diabetes or immunocompromising conditions such as solid organ transplants or hematologic malignancies are at risk for invasive rhinocerebral mucormycosis (See also [Chap. 213](#)). Patients present with low-grade fever, dull sinus pain, diplopia, decreased mental status, decreased ocular motion, chemosis, proptosis, dusky or necrotic nasal turbinates, and necrotic hard-palate lesions that respect the midline. Without rapid recognition and intervention, the process continues on an inexorable invasive course, with mortality rates of 50–85% or greater. Uncontrolled diabetes and increasing age are negative prognostic factors.

Acute Bacterial Endocarditis

This entity presents with a much more aggressive course than subacute endocarditis (See also [Chap. 123](#)). Bacteria such as *S. aureus*, *S. pneumoniae*, *L. monocytogenes*, *Haemophilus* species, and streptococci of groups A, B, and G attack native valves. Native-valve endocarditis caused by *S. aureus* (including MRSA strains) is increasing, particularly in health care settings. Mortality rates range from 10% to 40%. The host may have comorbid conditions such as underlying malignancy, diabetes mellitus, IV drug use, or alcoholism. The patient presents with fever, fatigue, and malaise <2 weeks after onset of infection. On physical examination, a changing murmur and congestive heart failure may be noted. Hemorrhagic macules on palms or soles (*Janeway lesions*) sometimes develop. Petechiae, Roth's spots, splinter hemorrhages, and splenomegaly are unusual. Rapid valvular destruction, particularly of the aortic valve, results in pulmonary edema and hypotension. Myocardial abscesses can form, eroding through the septum or into the conduction system and causing life-threatening arrhythmias or high-degree conduction block. Large friable vegetations can result in major arterial emboli, metastatic infection, or tissue infarction. Older patients with *S. aureus*

endocarditis are especially likely to present with nonspecific symptoms—a circumstance that delays diagnosis and worsens prognosis. Rapid intervention is crucial for a successful outcome.

Inhalational Anthrax

Inhalational anthrax, the most severe form of disease caused by *B. anthracis*, had not been reported in the United States for more than 25 years until the use of this organism as an agent of bioterrorism in 2001 (See also [Chap. S2](#)). Patients presented with malaise, fever, cough, nausea, drenching sweats, shortness of breath, and headache. Rhinorrhea was unusual. All patients had abnormal chest roentgenograms at presentation. Pulmonary infiltrates, mediastinal widening, and pleural effusions were the most common findings. Hemorrhagic meningitis was documented in 38% of these patients. Survival was more likely when antibiotics were given during the prodromal period and when multidrug regimens were used. In the absence of urgent intervention with antimicrobial agents and supportive care, inhalational anthrax progresses rapidly to hypotension, cyanosis, and death.

Viral Respiratory Tract Illness

Viral respiratory tract illnesses can cause severe disease; several new syndromes have been described in the past decade. For patients who present with a respiratory illness and a relevant exposure and travel history, these viral illnesses must be considered and appropriate infection control measures instituted in addition to supportive care.

Avian and Swine Influenza

Human cases of avian influenza have occurred primarily in Southeast Asia, particularly Vietnam (H5N1) and China (H7N9) (See also [Chap. 195](#)). Avian influenza should be considered in patients with severe respiratory tract illness, particularly if they have been exposed to poultry. Patients present with high fever, an influenza-like illness, and lower respiratory tract symptoms; this illness can progress rapidly to bilateral pneumonia, acute respiratory distress syndrome, multiorgan failure, and death. Younger age appears to be associated with a lower risk of complications. Early antiviral treatment with neuraminidase inhibitors should be initiated along with aggressive supportive measures. Unlike avian influenza, whose human-to-human transmission has so far been rare and has not been sustained, influenza caused by a novel swine-associated A/H1N1 virus has spread rapidly throughout the world; by 2012, 214 countries had diagnosed cases of influenza A/H1N1, with 18,449 deaths. Patients most at risk of severe disease are children <5 years of age, elderly persons, patients with underlying chronic conditions, and pregnant women. Obesity also has been identified as a risk factor for severe illness. Immunosuppression and co-infection with *S. aureus* at presentation are independent risk factors for increased mortality.

SARS and MERS

Severe acute respiratory syndrome (SARS) was identified in 2002 in China but has been diagnosed in several countries, primarily in Asia. Possible animal reservoirs include bats and civets. SARS is caused by a

coronavirus and is characterized by efficient human transmission but relatively low mortality. It spreads from person to person via droplets; “super-spreader” airborne events have occurred. The potential pandemic with SARS was controlled through identification and isolation of infected patients. A 3- to 7-day prodrome characterized by fever, malaise, headache, and myalgia can progress to nonproductive cough, dyspnea, and respiratory failure. The risk of contagion is low during the prodrome. Older patients and those with diabetes mellitus, chronic hepatitis B, and other comorbidities can have less favorable outcomes.

Middle East respiratory syndrome (MERS) is caused by a novel betacoronavirus and was first recognized in 2012 in Saudi Arabia. Human cases have been associated with direct and indirect contact with dromedary camels. Unlike SARS, MERS exhibits inefficient human transmission but carries a high mortality rate. As of 2015, 1180 cases had been confirmed, with 40% mortality. MERS ranges from asymptomatic infection to acute respiratory distress syndrome, multiorgan failure, and death. Elderly men with comorbidities appear to be at highest risk for poor outcomes. Despite little documented human-to-human transmission in the community, nosocomial infection must be prevented by adherence to strict infection control practices. MERS is currently a low-level public health threat and is likely to remain so unless the virus mutates and its transmissibility increases.

Hantavirus Pulmonary Syndrome

Hantavirus pulmonary syndrome has been documented in the United States since 1993 (primarily the southwestern states, west of the Mississippi River), Canada, and South America (See also [Chap. 204](#)). Most cases occur in rural areas and are associated with exposure to rodents. Patients present with a nonspecific viral prodrome of fever, malaise, myalgias, nausea, vomiting, and dizziness that may progress to pulmonary edema, respiratory failure, and death. Hantavirus pulmonary syndrome causes myocardial depression and increased pulmonary vascular permeability; therefore, careful fluid resuscitation and use of pressor agents are crucial. Aggressive cardiopulmonary support during the first few hours of illness can be life-saving in this high-mortality syndrome. The early onset of thrombocytopenia may help distinguish this syndrome from other febrile illnesses in an appropriate epidemiologic setting.

***Clostridium difficile* Infection**

C. difficile infection (CDI) is a toxin-mediated diarrheal syndrome that is strongly associated with prior antibiotic use. Proton-pump inhibitors have also been identified as a potential risk factor for the disease. Although most cases of CDI have occurred in the health care setting, community-onset CDI is increasing. Overall, community-onset cases occur in younger patients than nosocomial cases. Patients with community-onset CDI are less likely to have a history of antibiotic or proton-pump inhibitor use. CDI is associated with significant morbidity and mortality, particularly among older patients. The Centers for Disease Control and Prevention has reported that *C. difficile* infection is one of the top three health threats associated with antibiotic use.

SUMMARY

Acutely ill febrile patients with the syndromes discussed in this chapter require close observation, aggressive supportive measures, and—in most cases—admission to intensive care units. The most important task of the physician is to distinguish these patients from other infected febrile patients whose illness will not progress to fulminant disease. The alert physician must recognize the acute infectious disease emergency and then proceed with appropriate urgency.

FURTHER READING

Nicolasora N, Kaul DR: Infectious disease emergencies. Med Clin N Am 92:427, 2008. [[PubMed: 18298987](#)]

Seymour CW et al: Assessment of clinical criteria for sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 315:762, 2016. [[PubMed: 26903335](#)]

Simmons CP et al: Dengue. N Engl J Med 366:15, 2012.

Theilacker C et al: Overwhelming postsplenectomy infection: A prospective multicenter cohort study. Clin Infect Dis 62:871, 2016. [[PubMed: 26703862](#)]

Thigpen MC et al: Bacterial meningitis in the United States, 1998–2007. N Engl J Med 364:2016, 2011.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 118: Immunization Principles and Vaccine Use

Nancy Messonnier; Anne Schuchat; Lisa A. Jackson

INTRODUCTION

Few medical interventions of the past century can rival the effect that immunization has had on longevity, economic savings, and quality of life. Seventeen diseases are now preventable through vaccines routinely administered to children and adults in the United States (**Table 118-1**), and most vaccine-preventable diseases of childhood are at historically low levels (**Table 118-2**). Health care providers deliver the vast majority of vaccines in the United States in the course of providing routine health services and therefore play an integral role in the nation's public health system.

TABLE 118-1

Diseases Preventable with Vaccines Routinely Administered in the United States to Children and/or Adults

CONDITION	TARGET POPULATION(S) FOR ROUTINE USE
Pertussis	Children, adolescents, adults
Diphtheria	Children, adolescents, adults
Tetanus	Children, adolescents, adults
Poliomyelitis	Children
Measles	Children
Mumps	Children
Rubella, congenital rubella syndrome	Children
Hepatitis B	Children and high-risk adults
<i>Haemophilus influenzae</i> type b infection	Children
Hepatitis A	Children
Influenza	Children, adolescents, adults
Varicella	Children
Pneumococcal disease	Children, older adults
Meningococcal disease	Adolescents and high-risk adults
Rotavirus infection	Infants
Human papillomavirus infection, cervical and anogenital cancers	Adolescents and young adults
Zoster	Older adults

TABLE 118-2

Decline in Vaccine-Preventable Diseases in the United States Following Widespread Implementation of National Vaccine Recommendations

CONDITION	ANNUAL NO. OF PREVACCINE CASES (AVERAGE)	NO. OF CASES REPORTED IN 2016 ^a	REDUCTION (%) IN CASES AFTER WIDESPREAD VACCINATION
Smallpox	29,005	0	100
Diphtheria	21,053	0	100
Measles	530,217	69	>99
Mumps	162,344	5311	97
Pertussis	200,752	15,737	92
Polio (paralytic)	16,316	0	100
Rubella	47,745	5	>99
Congenital rubella syndrome	152	1	99
Tetanus	580	33	94
<i>Haemophilus influenzae</i> type b infection	20,000	22 ^b	>99
Hepatitis A	117,333	2500 ^c	98
Hepatitis B (acute)	66,232	19,200 ^c	71
Invasive pneumococcal infection: all ages	63,067	29,000 ^d	54
Varicella	4,085,120	126,639 ^e	97

^a
2016 reported cases unless otherwise specified.

^b
An additional 11 type b infections are estimated to have occurred among 222 reports of *H. influenzae* infection caused by unknown types among children <5 years of age.

^c
Data are from the CDC's Viral Hepatitis Surveillance, 2014.

^d
Data are from the CDC's Active Bacterial Core Surveillance 2015 Provisional Report.

^e
Data are from Morb Mortal Wkly Rep 65:1306, 2016 (2015 final data).

Source: Adapted from SW Roush et al: JAMA 298:2155, 2007 and Morb Mortal Wkly Rep 65:924, 2017.

VACCINE IMPACT

Direct and Indirect Effects

Immunizations against specific infectious diseases protect individuals against infection and thereby prevent symptomatic illnesses. Specific vaccines may blunt the severity of clinical illness (e.g., rotavirus vaccines and severe gastroenteritis) or reduce complications (e.g., zoster vaccines and postherpetic neuralgia). Some immunizations also reduce transmission of infectious disease agents from immunized people to others, thereby reducing the impact of infection spread. This indirect impact is known as *herd immunity*. The level of immunization in a population that is required to achieve indirect protection of unimmunized people varies substantially with the specific vaccine and disease.

Since childhood vaccines have become widely available in the United States, major declines in rates of vaccine-preventable diseases among both children and adults have become evident (Table 118-2). For example, vaccination of children <5 years of age against seven types of *Streptococcus pneumoniae* led to a >90% overall reduction in invasive disease caused by those types. Among children born during 1994–2013, a series of childhood vaccines targeting 13 vaccine-preventable diseases will prevent 322 million illnesses and 732,000 deaths over the course of their lifetimes and save \$1.38 trillion (U.S.).

Control, Elimination, and Eradication of Vaccine-Preventable Diseases

Immunization programs are associated with the goals of controlling, eliminating, or eradicating a disease. *Control* of a vaccine-preventable disease reduces poor illness outcomes and often limits the disruptive impacts associated with outbreaks of disease in communities, schools, and institutions. Control programs can also reduce absences from work for ill persons and for parents caring for sick children, decrease absences from school, and limit health care utilization associated with treatment visits.

Elimination of a disease is a more demanding goal than control, usually requiring the reduction to zero of cases in a defined geographic area but sometimes defined as reduction in the indigenous sustained transmission of an infection in a geographic area. As of 2016, the United States had eliminated indigenous transmission of measles, rubella, poliomyelitis, and diphtheria. Importation of pathogens from other parts of the world continues to be important, and public health efforts are intended to respond promptly to such cases in order to limit forward spread of the infectious agent.



Eradication of a disease is achieved when its elimination can be sustained without the need to continue interventions. The only vaccine-preventable disease of humans that has been globally eradicated thus far is smallpox. Although smallpox vaccine is no longer given routinely, the disease has not reemerged naturally because all chains of human transmission were interrupted through earlier vaccination efforts and humans were the only natural reservoir of the virus. Currently, a major health initiative is targeting the global eradication of polio. Sustained transmission of polio has been eliminated from most nations but has not yet been interrupted in Afghanistan and Pakistan. In 2016, after Nigeria completed 2 years without a wild polio case detected, three cases were identified in a region where vaccinators have been unable to reach hundreds of thousands of children because of insurgency and the virus is likely to have been circulating undetected. Detection of a case of disease that has been targeted for eradication or elimination is considered a sentinel event that could permit the infectious agent to become reestablished in the community or region. Therefore, such episodes must be promptly reported to public health authorities.

Outbreak Detection and Control

Clusters of cases of a vaccine-preventable disease detected in an institution, a medical practice, or a community may signal important changes in the pathogen, vaccine, or environment. Several factors can give rise to increases in vaccine-preventable disease, including (1) low rates of immunization that result in an accumulation of susceptible people (e.g., measles resurgence among vaccination abstainers); (2) changes in the infectious agent that permit it to escape vaccine-induced protection (e.g., non-vaccine-type pneumococci); (3) waning of vaccine-induced immunity (e.g., pertussis among adolescents and adults vaccinated in early childhood); and (4) point-source introductions of large inocula (e.g., food-borne exposure to hepatitis A virus). Reporting episodes of outbreak-prone diseases to public health authorities can facilitate recognition of clusters that require further interventions.

Public Health Reporting

Recognition of suspected cases of diseases targeted for elimination or eradication—along with other diseases that require urgent public health interventions, such as contact tracing, administration of chemo- or immunoprophylaxis, or epidemiologic investigation for common-source exposure—is typically associated with special reporting requirements. Many diseases against which vaccines are routinely used, including measles, pertussis, *Haemophilus influenzae* type b invasive disease, and varicella, are nationally notifiable. Clinicians and laboratory staff have a responsibility to report some vaccine-preventable disease occurrences to local or state public health authorities according to specific case-definition criteria. All providers should be aware of state or city disease-reporting requirements and the best ways to contact public health authorities. A prompt response to vaccine-preventable disease outbreaks can greatly enhance the effectiveness of control measures.

Global Considerations



Several international health initiatives currently focus on reducing vaccine-preventable diseases in regions throughout the world. The American Red Cross, the World Health Organization (WHO), the United Nations Foundation, the United Nations Children's Fund (UNICEF), and the Centers for Disease Control and Prevention (CDC) are partners in the Measles & Rubella Initiative, which targets reduction of worldwide measles deaths. During 2000–2014, global measles mortality rates declined by 78%—i.e., from an estimated 535,300 deaths in 2000 to 114,900 deaths in 2014. In 2015, the Americas became the first WHO region to be declared free of endemic transmission of rubella. Rotary International, UNICEF, the CDC, and the WHO are leading partners in the global eradication of polio, an endeavor that reduced the annual number of paralytic polio cases from 350,000 in 1988 to 74 in 2015. The GAVI Alliance and the Bill and Melinda Gates Foundation have brought substantial momentum to global efforts to reduce vaccine-preventable diseases, expanding on earlier efforts by the WHO, UNICEF, and governments in developed and developing countries.

Enhancing Immunization in Adults

Although immunization has become a centerpiece of routine pediatric medical visits, it has not been as well integrated into routine health care visits for adults. This chapter focuses on immunization principles and vaccine use in adults. Accumulating evidence suggests that immunization coverage can be increased through efforts directed at consumer-, provider-, institution-, and system-level factors. The literature suggests that the application of multiple strategies is more effective at raising coverage rates than is the use of any single strategy.

Recommendations for Adult Immunizations

The CDC's Advisory Committee on Immunization Practices (ACIP) is the main source of recommendations for administration of vaccines approved by the U.S. Food and Drug Administration (FDA) for use in children and adults in the U.S. civilian population. The ACIP is a federal advisory committee that consists of 15 voting members (experts in fields associated with immunization) appointed by the Secretary of the U.S. Department of Health and Human Services; 8 ex officio members representing federal agencies; and 30 nonvoting representatives of various liaison organizations, including major medical societies and managed-care organizations. The ACIP recommendations, which are available at www.cdc.gov/vaccines/hcp/acip-recs/, are harmonized to the greatest extent possible with vaccine recommendations made by other organizations, including the American College of Obstetricians and Gynecologists, the American Academy of Family Physicians, and the American College of Physicians.

Adult Immunization Schedules

Immunization schedules for adults in the United States are updated annually and can be found online (www.cdc.gov/vaccines/schedules/hcp/adult.html). In February, the schedules are published in *American Family Physician*, *Annals of Internal Medicine*, and *Morbidity and Mortality Weekly Report* (www.cdc.gov/mmwr). The adult immunization schedules for 2016 are summarized in **Fig. 118-1**. Additional information and specifications are contained in the footnotes to these schedules. In the time between annual publications, additions and changes to schedules are published in *Morbidity and Mortality Weekly Report*.

FIGURE 118-1

Recommended adult immunization schedules, United States, 2016. For complete statements by the Advisory Committee on Immunization Practices (ACIP), visit www.cdc.gov/vaccines/hcp/acip-recs/.

Recommended Adult Immunization Schedule—United States - 2016

Note: These recommendations must be read with the footnotes that follow containing number of doses, intervals between doses, and other important information.

Figure 1. Recommended immunization schedule for adults aged 19 years or older, by vaccine and age group¹

VACCINE ▼	AGE GROUP ►	19-21 years	22-26 years	27-49 years	50-59 years	60-64 years	≥ 65 years
Influenza ^{2,2}		1 dose annually					
Tetanus, diphtheria, pertussis (Td/Tdap) ^{3,3}		Substitute Tdap for Td once, then Td booster every 10 yrs					
Varicella ^{4,4}		2 doses					
Human papillomavirus (HPV) Female ^{5,5}		3 doses					
Human papillomavirus (HPV) Male ^{5,5}		3 doses					
Zoster ⁶						1 dose	
Measles, mumps, rubella (MMR) ^{7,7}		1 or 2 doses depending on indication					
Pneumococcal 13-valent conjugate (PCV13) ⁸		1 dose					
Pneumococcal 23-valent polysaccharide (PPSV23) ⁸		1 or 2 doses depending on indication					1 dose
Hepatitis A ⁹		2 or 3 doses depending on vaccine					
Hepatitis B ¹⁰		3 doses					
Meningococcal 4-valent conjugate (MenACWY) or polysaccharide (MPSV4) ^{11,11}		1 or more doses depending on indication					

Meningococcal B (MenB) ¹¹	2 or 3 doses depending on vaccine
<i>Haemophilus influenzae</i> type b (Hib) ^{*12}	1 or 3 doses depending on indication

^{*}Covered by the Vaccine Injury Compensation Program

Report all clinically significant postvaccination reactions to the Vaccine Adverse Event Reporting System (VAERS). Reporting forms and instructions on filing a VAERS report are available at www.vaers.hhs.gov or by telephone, 800-822-7967.

Information on how to file a Vaccine Injury Compensation Program claim is available at www.hrsa.gov/vaccinecompensation or by telephone, 800-338-2382. To file a claim for vaccine injury, contact the U.S. Court of Federal Claims, 717 Madison Place, N.W., Washington, D.C. 20005; telephone, 202-357-6400.

Additional information about the vaccines in this schedule, extent of available data, and contraindications for vaccination is also available at www.cdc.gov/vaccines or from the CDC-INFO Contact Center at 800-CDC-INFO (800-232-4636) in English and Spanish, 8:00 a.m. - 8:00 p.m. Eastern Time, Monday - Friday, excluding holidays.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

The recommendations in this schedule were approved by the Centers for Disease Control and Prevention's (CDC) Advisory Committee on Immunization Practices (ACIP), the American Academy of Family Physicians (AAFP), the American College of Physicians (ACP), the American College of Obstetricians and Gynecologists (ACOG) and the American College of Nurse-Midwives (ACNM).

Recommended for all persons who meet the age requirement, lack documentation of vaccination, or lack evidence of past infection; zoster vaccine is recommended regardless of past episode of zoster
 Recommended for persons with a risk factor (medical, occupational, lifestyle, or other indication)
 No recommendation

Figure 2. Vaccines that might be indicated for adults aged 19 years or older based on medical and other indications¹

VACCINE ▼	INDICATION ►	Pregnancy	Immuno-compromising conditions (excluding HIV infection) ^{4,6,7,8,13}	HIV infection (CD4+ count (cells/ μ L)) ^{4,6,7,8,13}	Men who have sex with men (MSM)	Kidney failure, end-stage renal disease, on hemodialysis	Heart disease, chronic lung disease, chronic alcoholism	Asplenia and persistent complement component deficiencies ^{9,11,12}	Chronic liver disease	Diabetes	Healthcare personnel
Influenza ^{*2}											
Tetanus, diphtheria, pertussis (Td/Tdap) ^{*3}		1 dose Tdap each pregnancy									
Varicella ^{*4}			Contraindicated								
Human papillomavirus (HPV) Female ^{*5}											
Human papillomavirus (HPV) Male ^{*5}											
Zoster ⁶			Contraindicated								
Measles, mumps, rubella (MMR) ^{*7}			Contraindicated								
Pneumococcal 13-valent conjugate (PCV13) ^{*8}											
Pneumococcal polysaccharide (PPSV23) ⁸											
Hepatitis A ⁹											
Hepatitis B ^{*10}											
Meningococcal 4-valent conjugate (MenACWY) or polysaccharide (MPSV4) ^{*11}											
Meningococcal B (MenB) ¹¹											
<i>Haemophilus influenzae</i> type b (Hib) ^{*12}											

^{*}Covered by the Vaccine Injury Compensation Program

Recommended for all persons who meet the age requirement, lack documentation of vaccination, or lack evidence of past infection; zoster vaccine is recommended regardless of past episode of zoster
 Recommended for persons with a risk factor (medical, occupational, lifestyle, or other indication)
 No recommendation
 Contraindicated



U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

These schedules indicate the recommended age groups and medical indications for which administration of currently licensed vaccines is commonly recommended for adults aged ≥ 19 years, as of February 2016. For all vaccines being recommended on the Adult Immunization Schedule, a vaccine series does not need to be restarted, regardless of the time that has elapsed between doses. Licensed combination vaccines may be used whenever any components of the combination are indicated and when the vaccine's other components are not contraindicated. For detailed recommendations on all vaccines, including those used primarily for travelers or that are issued during the year, consult the manufacturers' package inserts and the complete statements from the Advisory Committee on Immunization Practices (www.cdc.gov/vaccines/hcp/acip-recs/index.html). Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

IMMUNIZATION PRACTICE STANDARDS

Administering immunizations to adults involves a number of processes, such as deciding whom to vaccinate, assessing vaccine contraindications and precautions, providing vaccine information statements (VISs), ensuring appropriate storage and handling of vaccines, administering vaccines, and maintaining vaccine records. In addition, provider reporting of adverse events that follow vaccination is an essential component of the vaccine safety monitoring system. In 2014, the standards for adult immunization were revised to focus on vaccinating adults at every opportunity.

Deciding Whom to Vaccinate

Every effort should be made to ensure that adults receive all indicated vaccines as expeditiously as possible. When adults present for care, their immunization history should be assessed and recorded, and this information should be used to identify needed vaccinations according to the most current version of the adult immunization schedule. Decision-support tools incorporated into electronic health records can provide prompts for needed vaccinations. Standing orders, which are often used for routinely indicated vaccines (e.g., influenza and pneumococcal vaccines), permit a nurse or another approved licensed practitioner to administer vaccines without a specific physician order, thus lowering barriers to adult immunization.

Assessing Contraindications and Precautions

Before vaccination, all patients should be screened for contraindications and precautions. A *contraindication* is a condition that increases the risk of a serious adverse reaction to vaccination. A vaccine should not be administered when a contraindication is documented. For example, a history of an anaphylactic reaction to a dose of vaccine or to a vaccine component is a contraindication for further doses. A *precaution* is a condition that may increase the risk of an adverse event or that may compromise the ability of the vaccine to evoke immunity (e.g., administering measles vaccine to a person who has recently received a blood transfusion and may consequently have transient passive immunity to measles virus). Normally, a vaccine is not administered when a precaution is noted. However, situations may arise when the benefits of vaccination outweigh the estimated risk of an adverse event, and the provider may decide to vaccinate the patient despite the precaution.

In some cases, contraindications and precautions are temporary and may lead to mere deferral of vaccination until a later time. For example, moderate or severe acute illness with or without fever is generally considered a transient precaution to vaccination and results in postponement of vaccine administration until the acute phase has resolved; thus the superimposition of adverse effects of vaccination on the underlying illness and the mistaken attribution of a manifestation of the underlying illness to the vaccine are avoided. Contraindications and precautions to vaccines licensed in the United States for use in civilian adults are summarized in [Table 118-3](#). It is important to recognize conditions that are *not* contraindications in order not to miss opportunities for vaccination. For example, in most cases, mild acute illness (with or without fever), a history of a mild to moderate local reaction to a previous dose of the vaccine, and breast-feeding are not contraindications to vaccination.

Contraindications and Precautions for Commonly Used Vaccines in Adults

VACCINE FORMULATION	CONTRAINDICATIONS AND PRECAUTIONS
All vaccines	<p>Contraindication</p> <p>Severe allergic reaction (e.g., anaphylaxis) after a previous vaccine dose or to a vaccine component</p> <p>Precaution</p> <p>Moderate or severe acute illness with or without fever. Defer vaccination until illness resolves.</p>
Td	<p>Precautions</p> <p>GBS within 6 weeks after a previous dose of TT-containing vaccine</p> <p>History of Arthus-type hypersensitivity reactions after a previous dose of TD- or DT-containing vaccines (including MCV4). Defer vaccination until at least 10 years have elapsed since the last dose.</p> <p>History of severe allergic reaction to dry natural rubber (latex) (certain formulations; syringe; see text)</p>
Tdap	<p>Contraindication</p> <p>History of encephalopathy (e.g., coma or prolonged seizures) not attributable to another identifiable cause within 7 days of administration of a vaccine with pertussis components, such as DTaP or Tdap</p> <p>Precautions</p> <p>GBS within 6 weeks after a previous dose of TT-containing vaccine</p> <p>Progressive or unstable neurologic disorder, uncontrolled seizures, or progressive encephalopathy. Defer vaccination until a treatment regimen has been established and the condition has stabilized.</p> <p>History of Arthus-type hypersensitivity reactions after a previous dose of TT- or DT-containing vaccines (including MCV4). Defer vaccination until at least 10 years have elapsed since the last dose.</p> <p>History of severe allergic reaction to dry natural rubber (latex) (syringe; see text)</p>
HPV	<p>Contraindications</p> <p>History of immediate hypersensitivity to yeast (for Gardasil)</p> <p>History of severe allergic reaction to dry natural rubber (latex) (certain formulations; see text)</p> <p>Precaution</p> <p>Pregnancy (If a woman is found to be pregnant after initiation of the vaccination series, the remainder of the 3-dose regimen should be delayed until after completion of the pregnancy. If a vaccine dose has been administered during pregnancy, no intervention is needed. Exposure to Gardasil during pregnancy should be reported to Merck at 800-986-8999; exposure to Cervarix during pregnancy should be reported to GlaxoSmithKline at 888-452-9622.)</p>
MMR	<p>Contraindications</p> <p>History of immediate hypersensitivity reaction to gelatin^a or neomycin</p> <p>Pregnancy</p> <p>Known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection)</p> <p>Precautions</p> <p>Recent receipt (within 11 months) of antibody-containing blood product</p> <p>History of thrombocytopenia or thrombocytopenic purpura</p>
Varicella	<p>Contraindications</p> <p>Pregnancy</p> <p>Known severe immunodeficiency</p> <p>History of immediate hypersensitivity reaction to gelatin^a or neomycin</p> <p>Precaution</p> <p>Recent receipt (within 11 months) of antibody-containing blood product</p>

VACCINE FORMULATION	CONTRAINDICATIONS AND PRECAUTIONS
Influenza, inactivated, injectable	<p>Contraindication History of severe allergic reaction to dry natural rubber (latex) (certain formulations; see text)</p> <p>Precaution^b History of GBS within 6 weeks after a previous influenza vaccine dose</p>
Influenza, live attenuated nasal spray	<p>Contraindications^b Age ≥ 50 years Pregnancy Immunosuppression, including that caused by medications or by HIV infection; known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection) Certain chronic medical conditions, such as diabetes mellitus; chronic pulmonary disease (including asthma); chronic cardiovascular disease (except hypertension); renal, hepatic, neurologic/neuromuscular, hematologic, or metabolic disorders Close contact with severely immunosuppressed persons who require a protected environment, such as isolation in a bone marrow transplantation unit Close contact with persons with lesser degrees of immunosuppression (e.g., persons receiving chemotherapy or radiation therapy who are not being cared for in a protective environment; persons with HIV infection) is <i>not</i> a contraindication or a precaution. Health care personnel in neonatal intensive care units or oncology clinics may receive live attenuated influenza vaccine.</p> <p>Precautions History of GBS within 6 weeks of a previous influenza vaccine dose Receipt of specific antiviral agents (i.e., amantadine, rimantadine, zanamivir, or oseltamivir) within 48 h before vaccination</p>
Pneumococcal polysaccharide	None, other than those listed for all vaccines
Pneumococcal conjugate	None, other than those listed for all vaccines
Hepatitis A	<p>Contraindication History of severe allergic reaction to dry natural rubber (latex) (syringe; see text)</p>
Hepatitis B	<p>Contraindications History of immediate hypersensitivity to yeast History of severe allergic reaction to dry natural rubber (latex) (syringe; see text)</p>
Meningococcal conjugate	None, other than those listed for all vaccines
Meningococcal polysaccharide	<p>Contraindication History of severe allergic reaction to dry natural rubber (latex)</p>
Serogroup B meningococcal	<p>Contraindication History of severe allergic reaction to dry natural rubber (latex) (certain formulations; syringe; see text)</p>
Zoster	<p>Contraindications Pregnancy Known severe immunodeficiency History of immediate hypersensitivity reaction to gelatin^a or neomycin</p> <p>Precaution Receipt of specific antiviral agents (i.e., acyclovir, famciclovir, or valacyclovir) within 24 h before vaccination</p>

a

Extreme caution must be exercised in administering MMR, varicella, or zoster vaccine to persons with a history of anaphylactic reaction to gelatin or gelatin-containing products. Before administration, skin testing for sensitivity to gelatin can be considered. However, no specific protocols for this purpose have been published.

b

History of severe allergic reaction (e.g., anaphylaxis) to egg is a labeled contraindication to the use of inactivated influenza vaccine and live attenuated influenza vaccine. However, CDC's Advisory Committee on Immunization Practices recommends that any licensed, recommended, and appropriate inactivated influenza vaccine or recombinant influenza vaccine may be administered to persons with egg allergy of any severity (www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/flu.html).

Abbreviations: DT, diphtheria toxoid; DTaP, diphtheria, tetanus, and pertussis; GBS, Guillain-Barré syndrome; HPV, human papillomavirus; MCV4, quadrivalent meningococcal conjugate vaccine; MMR, measles, mumps, and rubella; Td, tetanus and diphtheria toxoids; Tdap, tetanus and diphtheria toxoids and acellular pertussis; TT, tetanus toxoid.

History of Immediate Hypersensitivity to a Vaccine Component

A severe allergic reaction (e.g., anaphylaxis) to a previous dose of a vaccine or to one of its components is a contraindication to vaccination. While most vaccines have many components, substances to which individuals are most likely to have had a severe allergic reaction include egg protein, gelatin, and yeast. In addition, although natural rubber (latex) is not a vaccine component, some vaccines are supplied in vials or syringes that contain natural rubber latex. These vaccines can be identified by the product insert and should not be administered to persons who report a severe (anaphylactic) allergy to latex unless the benefit of vaccination clearly outweighs the risk for a potential allergic reaction. The much more common local or contact hypersensitivity to latex, such as to medical gloves (which contain synthetic latex that is not linked to allergic reactions), is *not* a contraindication to administration of a vaccine supplied in a vial or syringe that contains natural rubber latex. Vaccines routinely indicated for adults that, as of February 2015, were sometimes supplied in a vial or syringe containing natural rubber include Havrix hepatitis A vaccine (syringe); Vaxna hepatitis A vaccine (vial and syringe); Engerix-B hepatitis B vaccine (syringe); Recombivax HB hepatitis B vaccine (vial); Cervarix HPV vaccine (syringe); Fluvirin, Agriflu (syringe), and Flucelvax (syringe) influenza vaccines; Adacel and Boostrix Tdap (tetanus and diphtheria toxoids and acellular pertussis) vaccines (syringe); Td (tetanus and diphtheria toxoids) vaccines (syringe); Twinrix hepatitis A and B vaccine (syringe); Menomune meningococcal polysaccharide vaccine (vial); and Bexsero meningococcal serogroup B vaccine (syringe).

Pregnancy

Live-virus vaccines are contraindicated during pregnancy because of the hypothetical risk that vaccine virus replication will cause congenital infection or have other adverse effects on the fetus. Most live-virus vaccines, including varicella vaccine, are not secreted in breast milk; therefore, breast-feeding is not a contraindication for live-virus or other vaccines. Pregnancy is not a contraindication to administration of inactivated vaccines, but most are avoided during pregnancy because relevant safety data are limited. Two inactivated vaccines, Tdap vaccine and inactivated influenza vaccine, are routinely recommended for pregnant women in the United States. Tdap vaccine is recommended during each pregnancy, regardless of prior vaccination status, in order to prevent pertussis in neonates. Annual influenza vaccination is recommended for all persons 6 months of age and older, regardless of pregnancy status. Some other inactivated vaccines, such as meningococcal vaccines, may be given to pregnant women in certain circumstances.

Immunosuppression

Live-virus vaccines elicit an immune response due to replication of the attenuated (weakened) vaccine virus that is contained by the recipient's immune system. In persons with compromised immune function, enhanced replication of vaccine viruses is possible and could lead to disseminated infection with the vaccine virus. For this reason, live-virus vaccines are contraindicated for persons with severe immunosuppression, the definition of which may vary with the vaccine. Severe immunosuppression may be caused by many disease conditions, including HIV infection and hematologic or generalized malignancy. In some of these conditions, all affected persons are severely immunocompromised. In others (e.g., HIV infection), the degree to which the immune system is compromised depends on the severity of the condition, which in turn depends on the stage of disease or treatment. For example, measles-mumps-rubella (MMR) vaccine may be given to HIV-infected persons who are not severely immunocompromised. Severe immunosuppression may also be due to therapy with immunosuppressive agents, including high-dose glucocorticoids. In this situation, the dose, duration, and route of administration may influence the degree of immunosuppression.

VACCINE INFORMATION STATEMENTS

A VIS is a one-page (two-sided) information sheet produced by the CDC that informs vaccine recipients (or their parents or legal representatives) about the benefits and risks of a vaccine. VISs are mandated by the National Childhood Vaccine Injury Act (NCVIA) of 1986 and—whether the vaccine

recipient is a child or an adult—must be provided for any vaccine covered by the Vaccine Injury Compensation Program. As of July 2016, vaccines that are covered by the NCVIA and that are licensed for use in adults include Td, Tdap, hepatitis A, hepatitis B, human [papillomavirus](#), inactivated influenza, live intranasal influenza, MMR, pneumococcal conjugate, meningococcal conjugate, serogroup B meningococcal, polio, and varicella vaccines. When combination vaccines for which no separate VIS exists are given (e.g., hepatitis A and B combination vaccine), all relevant VISs should be provided. VISs also exist for some vaccines not covered by the NCVIA, such as pneumococcal polysaccharide, Japanese encephalitis, rabies, herpes zoster, typhoid, anthrax, and yellow fever vaccines. The use of these VISs is encouraged but is not mandated.

All current VISs are available on the Internet at two websites: the CDC's Vaccines & Immunizations site (www.cdc.gov/vaccines/hcp/vis/) and the Immunization Action Coalition's site (www.immunize.org/vis/). (The latter site also includes translations of the VISs.) VISs from these sites can be downloaded and printed.

STORAGE AND HANDLING

Injectable vaccines are packaged in multidose vials, single-dose vials, or manufacturer-filled single-dose syringes. The live attenuated nasal-spray influenza vaccine is packaged in single-dose sprayers. Oral typhoid vaccine is packaged in capsules. Some vaccines, such as MMR, varicella, zoster, and meningococcal polysaccharide vaccines, come as lyophilized (freeze-dried) powders that must be reconstituted (i.e., mixed with a liquid diluent) before use. The lyophilized powder and the diluent come in separate vials. Diluents are not interchangeable but rather are specifically formulated for each type of vaccine; only the specific diluent provided by the manufacturer for each type of vaccine should be used. Once lyophilized vaccines have been reconstituted, their shelf-life is limited and they must be stored under appropriate temperature and light conditions. For example, varicella and zoster vaccines must be protected from light and administered within 30 min of reconstitution; MMR vaccine likewise must be protected from light but can be used up to 8 h after reconstitution. Single-dose vials of [meningococcal polysaccharide vaccine](#) must be used within 30 min of reconstitution, while multidose vials must be used within 35 days.

Vaccines are stored either at refrigerator temperature (2–8°C) or at freezer temperature (–15°C or colder). In general, inactivated vaccines (e.g., inactivated influenza, pneumococcal polysaccharide, and meningococcal conjugate vaccines) are stored at refrigerator temperature, while vials of lyophilized-powder live-virus vaccines (e.g., varicella, zoster, and MMR vaccines) are stored at freezer temperature. Diluents for lyophilized vaccines may be stored at refrigerator or room temperature. Live attenuated influenza vaccine—a live-virus liquid formulation administered by nasal spray—is stored at refrigerator temperature.

Vaccine storage and handling errors can result in the loss of vaccines worth millions of dollars, and administration of improperly stored vaccines may elicit inadequate immune responses in patients. To improve the standard of vaccine storage and handling practices, the CDC has published detailed guidance (available at www.cdc.gov/vaccines/hcp/admin/storage/toolkit/storage-handling-toolkit.pdf). For vaccine storage, the CDC recommends stand-alone units—i.e., self-contained units that either refrigerate or freeze but do not do both—as these units maintain the required temperatures better than combination refrigerator/freezer units. Dormitory-style combined refrigerator/freezer units should never be used for vaccine storage.

The temperature of refrigerators and freezers used for vaccine storage must be monitored and recorded at least twice each workday. Ideally, continuous thermometers that measure and record temperature all day and all night are used, and minimal and maximal temperatures are read and documented each workday. The CDC recommends the use of calibrated digital thermometers with a probe in thermal-buffered material; more detailed information on specifications of storage units and temperature-monitoring devices is provided at the link given above.

ADMINISTRATION OF VACCINES

Most parenteral vaccines recommended for routine administration to adults in the United States are given by either the IM or the SC route; one influenza vaccine formulation approved for use in adults 18–64 years of age is given intradermally. Live-virus vaccines such as varicella, zoster, and MMR are given SC. Most inactivated vaccines are given IM. The 23-valent pneumococcal polysaccharide vaccine may be given either IM or SC, but IM administration is preferred because it is associated with a lower risk of injection-site reactions.

Vaccines given to adults by the SC route are administered with a 5/8-inch needle into the upper outer-triceps area. Vaccines administered to adults by the IM route are injected into the deltoid muscle ([Fig. 118-2](#)) with a needle whose length should be selected on the basis of the recipient's sex and weight to ensure adequate penetration into the muscle. Current guidelines indicate that, for men and women weighing <152 lbs (<70 kg), a 1-inch needle is sufficient; for women weighing 152–200 lbs (70–90 kg) and men weighing 152–260 lbs (70–118 kg), a 1- to 1.5-inch needle is needed; and for women weighing >200 lbs (>90 kg) and men weighing >260 lbs (>118 kg), a 1.5-inch needle is required. Additional illustrations of vaccine injection locations and techniques may be found at www.immunize.org/catg.d/p2020a.pdf.

FIGURE 118-2

Technique for IM administration of vaccine. (Photo credit: James Gathany, Centers for Disease Control and Prevention; accessible at Public Health Image Library, www.cdc.gov. PHIL ID#9420.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Aspiration, the process of pulling back on the plunger of the syringe after skin penetration but prior to injection, is not necessary because no large blood vessels are present at the recommended vaccine injection sites.

Multiple vaccines can be administered at the same visit; indeed, administration of all needed vaccines at one visit is encouraged. Studies have shown that vaccines are as effective when administered simultaneously as they are individually, and simultaneous administration of multiple vaccines is not associated with an increased risk of adverse effects. If more than one vaccine must be administered in the same limb, the injection sites should be separated by 1–2 inches so that any local reactions can be differentiated. If a vaccine and an **immune globulin** preparation are administered simultaneously (e.g., Td vaccine and tetanus **immune globulin**), a separate anatomic site should be used for each injection.

For certain vaccines (e.g., HPV vaccine and hepatitis B vaccine), multiple doses are required for an adequate and persistent antibody response. The recommended vaccination schedule specifies the interval between doses. Many adults who receive the first dose in a multiple-dose vaccine series do not complete the series or do not receive subsequent doses within the recommended interval; this lack of adherence to protocol compromises vaccine efficacy and/or the duration of protection. Providers should implement recall systems that will prompt patients to return for subsequent doses in a vaccination series at the appropriate intervals. With the exception of oral typhoid vaccination, an interruption in the schedule does not require restarting of the entire series or the addition of extra doses.

Syncope may follow vaccination, especially in adolescents and young adults. Serious injuries, including skull fracture and cerebral hemorrhage, have occurred. Adolescents and adults should be seated or lying down during vaccination. The majority of reported syncope episodes after vaccination occur within 15 min. The ACIP recommends that vaccine providers strongly consider observing patients, particularly adolescents, with patients seated or lying down for 15 min after vaccination. If syncope develops, patients should be observed until the symptoms resolve.

Anaphylaxis is a rare complication of vaccination. All facilities providing immunizations should have an emergency kit containing aqueous **epinephrine** for administration in the event of a systemic anaphylactic reaction.

MAINTENANCE OF VACCINE RECORDS

All vaccines administered should be fully documented in the patient's permanent medical record. Documentation should include the date of administration, the name or common abbreviation of the vaccine, the vaccine lot number and manufacturer, the administration site, the VIS edition, the date the VIS was provided, and the name, address, and title of the person who administered the vaccine. Increasing use of two-dimensional bar codes on vaccine vials and syringes that can be scanned for data entry into compatible electronic medical records and immunization information systems may facilitate more complete and accurate recording of required information.

VACCINE SAFETY MONITORING AND ADVERSE EVENT REPORTING

Prelicensure Evaluations of Vaccine Safety

Before vaccines are licensed by the FDA, they are evaluated in clinical trials with volunteers. These trials are conducted in three progressive phases. Phase 1 trials are small, usually involving fewer than 100 volunteers. Their purposes are to provide a basic evaluation of safety and to identify

common adverse events. Phase 2 trials, which are larger and may involve several hundred participants, collect additional information on safety and are usually designed to evaluate immunogenicity as well. Data gained from phase 2 trials can be used to determine the composition of the vaccine, the number of doses required, and a profile of common adverse events. Vaccines that appear promising are evaluated in phase 3 trials, which typically involve several hundred to several thousand volunteers and are generally designed to demonstrate vaccine efficacy and provide additional information on vaccine safety.

Postlicensure Monitoring of Vaccine Safety

After licensure, a vaccine's safety is assessed by several mechanisms. The NCIA of 1986 requires health care providers to report certain adverse events that follow vaccination. As a mechanism for that reporting, the Vaccine Adverse Event Reporting System (VAERS) was established in 1990 and is jointly managed by the CDC and the FDA. This safety surveillance system collects reports of adverse events associated with vaccines currently licensed in the United States. *Adverse events* are defined as untoward events that occur after immunization and that might be caused by the vaccine product or vaccination process. While the VAERS was established in response to the NCIA, any adverse event following vaccination—whether in a child or an adult, and whether or not it is believed to have actually been caused by vaccination—may be reported through the VAERS. The adverse events that health care providers are required to report are listed in the reportable-events table on the VAERS website at vaers.hhs.gov/reportable.htm. During 2011–2014, approximately 30,000 VAERS reports were filed annually, with ~7% reporting serious events resulting in hospitalization, life-threatening illness, disability, or death.

Anyone can file a VAERS report, including health care providers, manufacturers, and vaccine recipients or their parents or guardians. VAERS reports may be submitted online (<https://vaers.hhs.gov/reportevent.html>) or by completing a paper form requested by email (info@vaers.org) or phone (800-822-7967). The VAERS form asks for the following information: the type of vaccine received; the timing of vaccination; the time of onset of the adverse event; and the recipient's current illnesses or medications, history of adverse events following vaccination, and demographic characteristics (e.g., age and sex). This information is entered into a database. The individual who reported the adverse event then receives a confirmation letter by mail with a VAERS identification number that can be used if additional information is submitted later. In selected cases of serious adverse reaction, the patient's recovery status may be followed up at 60 days and 1 year after vaccination. The FDA and the CDC have access to VAERS data and use this information to monitor vaccine safety and conduct research studies. VAERS data (minus personal information) are also available to the public.

While the VAERS provides useful information on vaccine safety, this passive reporting system has important limitations. One is that events following vaccination are merely reported; the system cannot assess whether a given type of event occurs more often than expected after vaccination. A second is that event reporting is incomplete and is biased toward events that are believed to be more likely to be due to vaccination and that occur relatively soon after vaccination. To obtain more systematic information on adverse events occurring in both vaccinated and unvaccinated persons, the Vaccine Safety Datalink project was initiated in 1991. Directed by the CDC, this project includes nine managed-care organizations in the United States; member databases include information on immunizations, medical conditions, demographics, laboratory results, and medication prescriptions. The Department of Defense oversees a similar system monitoring the safety of immunizations among active-duty military personnel. In addition, postlicensure evaluations of vaccine safety may be conducted by the vaccine manufacturer. In fact, such evaluations are often required by the FDA as a condition of vaccine licensure.

CONSUMER ACCESS TO AND DEMAND FOR IMMUNIZATION

By removing barriers to the consumer or patient, providers and health care institutions can improve vaccine use. Financial barriers have traditionally been important constraints, particularly among uninsured adults. Even for insured adults, out-of-pocket costs associated with newer, more expensive adult vaccines (e.g., zoster vaccine) are an obstacle to be overcome. After influenza vaccine was included by Medicare for all beneficiaries in 1993, coverage among persons ≥65 years of age doubled (from ~30% in 1989 to >60% in 1997). Other strategies that enhance patients' access to vaccination include extended office hours (e.g., evening and weekend hours) and scheduled vaccination-only clinics where waiting times are reduced. Provision of vaccines outside the "medical home" (e.g., through occupational clinics, universities, pharmacies, and retail settings) can expand access for adults who do not make medical visits frequently. Increasing proportions of adults are being vaccinated in these settings.

Health promotion efforts aimed at increasing the demand for immunization are common. Direct-to-consumer advertising by pharmaceutical companies has been used for some newer adolescent and adult vaccines. Efforts to raise consumer demand for vaccines have not increased immunization rates unless implemented in conjunction with other strategies that target strengthening of provider practices or reduction of consumer barriers. Attitudes and beliefs related to vaccination can be considerable impediments to consumer demand. Many adults view vaccines as important for children but are less familiar with vaccinations targeting disease prevention in adults. Several vaccines are recommended for adults with certain medical risk factors, but self-identification as a high-risk individual is relatively rare. Communication research suggests that adults are motivated to get vaccines to protect their own health and many would get vaccinated to protect loved ones. Adults with chronic

conditions are more likely to be aware that they need to protect their own health. Some vaccines are explicitly recommended for persons at relatively low risk of serious complications, with the goal of reducing the risk of transmission to higher-risk contacts. For example, for protection of newborns, vaccinations against influenza and pertussis are recommended for pregnant women.

STRATEGIES FOR PROVIDERS AND HEALTH CARE FACILITIES

Recommendation from the Provider

Health care providers can have great influence on patients with regard to immunization. A recommendation from a doctor or nurse carries more weight than do recommendations from professional societies or endorsements by celebrities. Providers should be well informed about vaccine risks and benefits so that they can address patients' common concerns. The CDC, the American College of Physicians, and the American Academy of Family Physicians review and update the schedule for adult immunization on an annual basis and have developed educational materials to facilitate provider–patient discussions about vaccination (www.cdc.gov/vaccines/hcp.htm).

System Supports

Medical offices can incorporate a variety of methods to ensure that providers consistently offer specific immunizations to patients with indications for specific vaccines. Decision-support tools have been incorporated into some electronic health records to alert the provider when specific vaccines are indicated. Manual or automated reminders and standing orders have been discussed (see “[Deciding Whom to Vaccinate](#),” above) and have consistently improved vaccination coverage in both office and hospital settings. Most clinicians' estimates of their own performance diverge from objective measurements of their patients' immunization coverage; quantitative assessment and feedback have been shown in pediatric and adolescent practices to increase immunization performance significantly. Some health plans have instituted incentives for providers with high rates of immunization coverage. Specialty providers, including obstetrician–gynecologists, may be the only providers serving some high-risk patients with indications for selected vaccines (e.g., Tdap, influenza, or pneumococcal polysaccharide vaccine).

Immunization Requirements

Vaccination against selected communicable diseases is required for attendance at many universities and colleges as well as for service in the U.S. military or in some occupational settings (e.g., child care, laboratory, veterinary, and health care). Immunizations are recommended and sometimes required for travel to certain countries ([Chap. 119](#)).

Vaccination of Health Care Staff

A particular area of focus for medical settings is vaccination of health care workers, including those with and without direct patient-care responsibilities. The Joint Commission (which accredits health care organizations), the CDC's Healthcare Infection Control Practices Advisory Committee, and the ACIP all recommend influenza vaccination of all health care personnel; recommendations also focus on requiring documentation of declination for providers who do not accept annual influenza vaccination. As part of their participation in the Centers for Medicare and Medicaid Services' Hospital Inpatient Quality Reporting program, acute-care hospitals are required to report the proportion of their health care personnel who have received seasonal influenza vaccine. Some institutions and jurisdictions have added mandates on influenza vaccination of health care workers and have expanded on earlier requirements related to vaccination or proof of immunity for hepatitis B, measles, mumps, rubella, and varicella.

VACCINATION IN NONMEDICAL SETTINGS

Receipt of vaccination in medical offices is most frequent among young children and adults ≥ 65 years of age. Patients in these age groups make more office visits and are more likely to receive care in a consistent “medical home” than are older children, adolescents, and nonelderly adults. Vaccination outside the medical home can expand access to those whose health care visits are limited and reduce the burden on busy clinical practices. In some locations, financial constraints related to inventory and storage requirements have led providers to stock few or no vaccines. Outside private office and hospital settings, vaccination may also occur at health department venues, workplaces, retail sites (including pharmacies and supermarkets), and schools or colleges.

When vaccines are given in nonmedical settings, it remains important for standards of immunization practice to be followed. Consumers should be provided with information on how to report adverse events (e.g., via provision of a VIS), and procedures should ensure that documentation of vaccine administration is forwarded to the primary care provider and the state or city public health immunization registry. Detailed documentation may be required for employment, school attendance, and travel. Personalized health records can help consumers keep track of their immunizations, and some occupational health clinics have incorporated automated immunization reports that help employees stay up-to-date with

recommended vaccinations. Some pharmacy chain establishments are using automated systems to report immunization information to the state or local immunization information system.

PERFORMANCE MONITORING

Tracking of immunization coverage at national, state, institution, and practice levels can yield feedback to practitioners and programs and facilitate quality improvement. Healthcare Effectiveness Data and Information Set (HEDIS) measures related to adult immunization facilitate comparison of health plans. The CDC's National Immunization Survey and National Health Interview Survey provide selected information on immunization coverage among adults and track progress toward achievement of Healthy People 2020 targets for immunization coverage. Influenza and pneumococcal vaccine coverage rates have been higher among persons ≥ 65 years of age (60–70%) than among high-risk 18- to 64-year-olds. Figures on state-specific immunization coverage with pneumococcal polysaccharide and influenza vaccines (as measured through the CDC's Behavioral Risk Factor Surveillance System) reveal substantial geographic variation in coverage. There are persistent disparities in adult immunization coverage rates between whites and racial and ethnic minorities. In contrast, racial and economic disparities in immunization of young children have been dramatically reduced during the past 20 years. Much of this progress is attributed to the Vaccines for Children Program, which since 1994 has entitled uninsured children to receive free vaccines.

FUTURE TRENDS

Although most vaccines developed in the twentieth century targeted common acute infectious diseases of childhood, more recently developed vaccines prevent chronic conditions prevalent among adults. Hepatitis B vaccine prevents hepatitis B-related cirrhosis and hepatocellular carcinoma, and HPV vaccine prevents some types of cervical cancer, genital warts, and anogenital cancers and may also prevent some oropharyngeal cancers. A new herpes zoster subunit vaccine that was licensed in 2017 should substantially improve protection against zoster and postherpetic neuralgia. New targets of vaccine development and research may further broaden the definition of vaccine-preventable disease. Research is ongoing on vaccines to prevent insulin-dependent diabetes mellitus, nicotine addiction, and Alzheimer's disease. Expanding strategies for vaccine development are incorporating molecular approaches such as DNA, vector, and peptide vaccines. New technologies, such as the use of transdermal and other needle-less routes of administration, are being applied to vaccine delivery.

FURTHER READING

Centers for Disease Control and Prevention: *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 13th ed. J Hamborsky et al (eds). Washington DC, Public Health Foundation, 2015.

McNeil MM et al: The vaccine safety datalink: Successes and challenges monitoring vaccine safety. *Vaccine* 32:5390, 2014. [[PubMed: 25108215](#)]

NATIONAL VACCINE ADVISORY COMMITTEE: Recommendations from the National Vaccine Advisory Committee: Standards for Adult Immunization Practice. *Public Health Rep* 129:115, 2014.

Plotkin SA et al (eds): *Plotkin's Vaccines*, 7th ed. Philadelphia, Elsevier, 2017.

Schuchat A, Bell BP: Monitoring the impact of vaccines post-licensure: New challenges, new opportunities. *Expert Rev Vaccines* 7:437, 2008. [[PubMed: 18444891](#)]

Whitney CW et al: Benefits from immunization during the Vaccines for Children Program era—United States, 1994–2013. *MMWR* 63:352, 2014. [[PubMed: 24759657](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is 178.250.250.21

Access Provided by: National Medical Library

Silverchair

Chapter 121: Pneumonia

Lionel A. Mandell; Richard Wunderink

DEFINITION

Pneumonia is an infection of the pulmonary parenchyma. Despite being the cause of significant morbidity and mortality, it is often misdiagnosed, mistreated, and underestimated. Pneumonia historically was typically classified as community-acquired (CAP), hospital-acquired (HAP), or ventilator-associated (VAP). A fourth category, health care–associated pneumonia (HCAP), was introduced recently. This category was meant to encompass those cases of CAP that were caused by multidrug-resistant (MDR) pathogens typically associated with HAP. Unfortunately, the original definitions appear to have been overly sensitive, resulting in the treatment of a high proportion of patients who had community-onset pneumonia with broad-spectrum antibiotics consistent with HAP treatment. Retrospective studies have actually suggested a worse outcome when broad-spectrum antibiotics were used in these cases.

Rather than relying on a predefined subset or category of pneumonia cases, it is likely to be of greater value to assess each case individually on the basis of risk factors for infection with an MDR organism. Rather than originating in primary pneumonia research, the original HCAP definition was modified from a study of health care–associated bacteremia. Recent studies have more closely identified patients at risk for pathogens resistant to the antibiotics usually used; have defined risk factors for infection with methicillin-resistant *Staphylococcus aureus* (MRSA) independent of other MDR pathogens; and have found that at least two, if not three, risk factors are required before the probability of drug-resistant pathogens is sufficient to influence initial empirical broad-spectrum antibiotic therapy. These risk factors are listed in [Table 121-1](#).

TABLE 121-1

Risk Factors for Pathogens Resistant to Usual Therapy for Community-Acquired Pneumonia^a

MULTIDRUG-RESISTANT GRAM-NEGATIVE BACTERIA AND MRSA	NOSOCOMIAL MRSA	COMMUNITY-ACQUIRED MRSA
Hospitalization ≥ 2 days in previous 90 days Use of antibiotics in previous 90 days Immunosuppression Nonambulatory status Tube feedings Gastric acid suppression Severe COPD or bronchiectasis ^b	Hospitalization ≥ 2 days in previous 90 days Use of antibiotics in previous 90 days Chronic hemodialysis in previous 30 days Documented prior MRSA colonization Congestive heart failure Gastric acid suppression	Cavitary infiltrate or necrosis Gross hemoptysis Neutropenia Erythematous rash Concurrent influenza Young, previously healthy status Summer-month onset

^aCephalosporin/macrolide or respiratory fluoroquinolone. ^bRisk for *Pseudomonas aeruginosa* infection.

Abbreviations: COPD, chronic obstructive pulmonary disease; MRSA, methicillin-resistant *Staphylococcus aureus*.

This chapter deals with pneumonia in patients who are not considered to be immunocompromised.

Pneumonia in severely immunocompromised patients, some of whom overlap with the groups of patients considered in this chapter, warrants separate discussion (see Chaps. 70, 138, and 197).

PATHOPHYSIOLOGY

Pneumonia results from the proliferation of microbial pathogens at the alveolar level and the host's response to those pathogens. Microorganisms gain access to the lower respiratory tract in several ways. The most common is by aspiration from the oropharynx. Small-volume aspiration occurs frequently during sleep (especially in the elderly) and in patients with decreased levels of consciousness. Rarely, pneumonia occurs via hematogenous spread (e.g., from tricuspid endocarditis) or by contiguous extension from an infected pleural or mediastinal space.

Mechanical factors are critically important in host defense. The hairs and turbinates of the nares capture larger inhaled particles before they reach the lower respiratory tract. The branching architecture of the tracheobronchial tree traps microbes on the airway lining, where mucociliary clearance and local antibacterial factors either clear or kill the potential pathogen. The gag and cough reflexes offer critical protection from aspiration. In addition, the normal flora adhering to mucosal cells of the oropharynx, whose

components are remarkably constant, prevents pathogenic bacteria from binding and thereby decreases the risk of pneumonia.

When these barriers are overcome or when microorganisms are small enough to be inhaled to the alveolar level, resident alveolar macrophages are extremely efficient at clearing and killing pathogens. Macrophages are assisted by proteins that are produced by the alveolar epithelial cells (e.g., surfactant proteins A and D) and that have intrinsic opsonizing properties or antibacterial or antiviral activity. Once engulfed by the macrophage, the pathogens—even if they are not killed—are eliminated via either the mucociliary elevator or the lymphatics and no longer represent an infectious challenge. Only when the capacity of the alveolar macrophages to ingest or kill the microorganisms is exceeded does clinical pneumonia become manifest. In that situation, the alveolar macrophages initiate the inflammatory response to bolster lower respiratory tract defenses. The host inflammatory response, rather than proliferation of microorganisms, triggers the clinical syndrome of pneumonia. The release of inflammatory mediators, such as interleukin 1 and tumor necrosis factor, results in fever. Chemokines, such as interleukin 8 and granulocyte colony-stimulating factor, stimulate the release of neutrophils and their attraction to the lung, producing both peripheral leukocytosis and increased purulent secretions. Inflammatory mediators released by macrophages and the newly recruited neutrophils create an alveolar capillary leak equivalent to that seen in acute respiratory distress syndrome, although in pneumonia this leak is localized (at least initially). Even erythrocytes can cross the alveolar–capillary membrane, with consequent hemoptysis. The capillary leak results in a radiographic infiltrate and rales detectable on auscultation, and hypoxemia results from alveolar filling. Moreover, some bacterial pathogens appear to interfere with the hypoxemic vasoconstriction that would normally occur with fluid-filled alveoli, and this interference can result in severe hypoxemia. Increased respiratory drive in the systemic inflammatory response syndrome ([Chap. 297](#)) leads to respiratory alkalosis. Decreased compliance due to capillary leak, hypoxemia, increased respiratory drive, increased secretions, and occasionally infection-related bronchospasm all lead to dyspnea. If severe enough, the changes in lung mechanics secondary to reductions in lung volume and compliance and the intrapulmonary shunting of blood may cause respiratory failure and death.

The presence of a normal alveolar microbiota raises the possibility of an alternative pathway for development of pneumonia. This microbiota is similar to the oropharyngeal microbiota; both are predominantly gram-positive in contrast to the gram-negative milieu of the normal gastrointestinal microbiota. Rather than invasion of a sterile lower respiratory tract by pathogens to cause pneumonia, alterations in host defense may allow overgrowth of one or more components of the normal bacterial flora. The fact that many CAP pathogens are components of the normal alveolar microbiota supports this alternative-pathogenesis model. The two most likely sources of an altered alveolar microbiota are viral upper respiratory tract infections for CAP and antibiotic therapy for HAP/VAP.

PATHOLOGY

Classic pneumonia evolves through a series of pathologic changes. The initial phase is one of *edema*, with the presence of a proteinaceous exudate—and often of bacteria—in the alveoli. This phase is rarely evident in

clinical or autopsy specimens because of the rapid transition to the *red hepatization* phase. The presence of erythrocytes in the cellular intra-alveolar exudate gives this second stage its name, but neutrophil influx is more important with regard to host defense. Bacteria are occasionally seen in pathologic specimens collected during this phase. In the third phase, *gray hepatization*, no new erythrocytes are extravasating, and those already present have been lysed and degraded. The neutrophil is the predominant cell, fibrin deposition is abundant, and bacteria have disappeared. This phase corresponds with successful containment of the infection and improvement in gas exchange. In the final phase, *resolution*, the macrophage reappears as the dominant cell type in the alveolar space, and the debris of neutrophils, bacteria, and fibrin has been cleared, as has the inflammatory response.

This pattern has been described best for lobar pneumococcal pneumonia and may not apply to pneumonia of all etiologies, especially viral or *Pneumocystis* pneumonia. In VAP, respiratory bronchiolitis may precede the development of a radiologically apparent infiltrate. Because of the microaspiration mechanism, a bronchopneumonia pattern is most common in nosocomial pneumonias, whereas a lobar pattern is more common in bacterial CAP. Despite the radiographic appearance, viral and *Pneumocystis* pneumonias represent alveolar rather than interstitial processes.

COMMUNITY-ACQUIRED PNEUMONIA

ETIOLOGY

The extensive list of potential etiologic agents in CAP includes bacteria, fungi, viruses, and protozoa. Newly identified pathogens include metapneumoviruses, the coronaviruses responsible for severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), and community-acquired strains of MRSA. Most cases of CAP, however, are caused by relatively few pathogens ([Table 121-2](#)). Although *Streptococcus pneumoniae* is most common, other organisms must also be considered in light of the patient's risk factors and severity of illness. Separation of potential agents into "typical" bacterial pathogens or "atypical" organisms may be helpful. The former category includes *S. pneumoniae*, *Haemophilus influenzae*, and (in selected patients) *S. aureus* and gram-negative bacilli such as *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*. The "atypical" organisms include *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, and *Legionella* species as well as respiratory viruses such as influenza viruses, adenoviruses, human metapneumovirus, and respiratory syncytial viruses. Overall, with the increasing use of pneumococcal vaccine, the incidence of pneumococcal pneumonia appears to be decreasing. Cases due to *M. pneumoniae* and *C. pneumoniae*, however, appear to be increasing in incidence, especially among young adults. Viruses may be responsible for a large proportion of CAP cases that require hospital admission, even in adults. Polymerase chain reaction (PCR)-based testing shows that viruses may be present in 20–30% of healthy adults and in the same percentage of patients with pneumonia, including those who are severely ill. The most common of these viruses are influenza, parainfluenza, and respiratory syncytial viruses. Whether they are etiologic pathogens, co-pathogens, or simply colonizers cannot always be determined. Atypical organisms cannot be cultured on standard media or seen on Gram's stain. The frequency and importance of atypical pathogens have significant implications for therapy. They are intrinsically resistant to all β -lactam

agents and must be treated with a macrolide, a fluoroquinolone, or a tetracycline. In the ~10–15% of CAP cases that are polymicrobial, the etiology usually includes a combination of typical and atypical pathogens.

TABLE 121-2

Microbial Causes of Community-Acquired Pneumonia, by Site of Care

OUTPATIENTS	HOSPITALIZED PATIENTS	
	NON-ICU	ICU
<i>Streptococcus pneumoniae</i>	<i>S. pneumoniae</i>	<i>S. pneumoniae</i>
<i>Mycoplasma pneumoniae</i>	<i>M. pneumoniae</i>	<i>Staphylococcus aureus</i>
<i>Haemophilus influenzae</i>	<i>Chlamydia pneumoniae</i>	<i>Legionella</i> spp.
<i>C. pneumoniae</i>	<i>H. influenzae</i>	Gram-negative bacilli
Respiratory viruses ^a	<i>Legionella</i> spp.	<i>H. influenzae</i>
	Respiratory viruses ^a	Respiratory viruses

a

Influenza A and B viruses, human metapneumovirus, adenoviruses, respiratory syncytial viruses, parainfluenza viruses.

Abbreviation: ICU, intensive care unit.

Anaerobes play a significant role only when an episode of aspiration has occurred days to weeks before presentation for pneumonia. The combination of an unprotected airway (e.g., in patients with alcohol or drug overdose or a seizure disorder) and significant gingivitis constitutes the major risk factor. Anaerobic pneumonias are often complicated by abscess formation and by significant empyemas or parapneumonic effusions.

S. aureus pneumonia is well known to complicate influenza infection. However, MRSA has been reported as a primary etiologic agent of CAP. While this entity is still relatively uncommon, clinicians must be aware of its potentially serious consequences, such as necrotizing pneumonia. Two important developments have led to this problem: the spread of MRSA from the hospital setting to the community and the emergence of genetically distinct strains of MRSA in the community. The community-acquired MRSA (CA-MRSA) strains may infect healthy individuals with no association with health care.

Unfortunately, despite a careful history and physical examination as well as routine radiographic studies, the causative pathogen in a case of CAP is difficult to predict with any degree of certainty; in more than one-half of cases, a specific etiology is never determined. Nevertheless, epidemiologic and risk factors may suggest the involvement of certain pathogens (**Table 121-3**).

TABLE 121-3

Epidemiologic Factors Suggesting Possible Causes of Community-Acquired Pneumonia

FACTOR	POSSIBLE PATHOGEN(S)
Alcoholism	<i>Streptococcus pneumoniae</i> , oral anaerobes, <i>Klebsiella pneumoniae</i> , <i>Acinetobacter</i> spp., <i>Mycobacterium tuberculosis</i>
COPD and/or smoking	<i>Haemophilus influenzae</i> , <i>Pseudomonas aeruginosa</i> , <i>Legionella</i> spp., <i>S. pneumoniae</i> , <i>Moraxella catarrhalis</i> , <i>Chlamydia pneumoniae</i>
Structural lung disease (e.g., bronchiectasis)	<i>P. aeruginosa</i> , <i>Burkholderia cepacia</i> , <i>Staphylococcus aureus</i>
Dementia, stroke, decreased level of consciousness	Oral anaerobes, gram-negative enteric bacteria
Lung abscess	CA-MRSA, oral anaerobes, endemic fungi, <i>M. tuberculosis</i> , atypical mycobacteria
Travel to Ohio or St. Lawrence river valley	<i>Histoplasma capsulatum</i>
Travel to southwestern United States	Hantavirus, <i>Coccidioides</i> spp.
Travel to Southeast Asia	<i>Burkholderia pseudomallei</i> , avian influenza virus
Stay in hotel or on cruise ship in previous 2 weeks	<i>Legionella</i> spp.
Local influenza activity	Influenza virus, <i>S. pneumoniae</i> , <i>S. aureus</i>
Exposure to bats or birds	<i>H. capsulatum</i>
Exposure to birds	<i>Chlamydia psittaci</i>
Exposure to rabbits	<i>Francisella tularensis</i>
Exposure to sheep, goats, parturient cats	<i>Coxiella burnetii</i>

Abbreviations: CA-MRSA, community-acquired methicillin-resistant *Staphylococcus aureus*; COPD, chronic obstructive pulmonary disease.

EPIDEMIOLOGY

More than 5 million CAP cases occur annually in the United States. Along with influenza, CAP is the eighth leading cause of death in this country. Usually, 80% of the affected patients are treated as outpatients and 20% as inpatients. The mortality rate among outpatients is usually <5%, whereas among hospitalized patients the rate can range from ~12% to 40%, depending on whether treatment is provided in or outside of the intensive care unit (ICU). In the United States, CAP is the number one cause of death from infection among patients >65 years of age. Further compounding its impact is the fact that 18% of hospitalized CAP patients are readmitted within 1 month of discharge. CAP results in more than 1.2 million hospitalizations and more than 55,000 deaths annually. The overall yearly cost associated with CAP is estimated at \$17 billion. The incidence rates are highest at the extremes of age. The overall annual rate in the United States is 12 cases/1000 persons, but the figure increases to 12–18/1000 among children <4 years of age and to 20/1000 among persons >60 years of age.

The risk factors for CAP in general and for pneumococcal pneumonia in particular have implications for treatment regimens. Risk factors for CAP include alcoholism, asthma, immunosuppression, institutionalization, and an age of ≥ 70 years. In the elderly, factors such as decreased cough and gag reflexes as well as reduced antibody and Toll-like receptor responses increase the likelihood of pneumonia. Risk factors for pneumococcal pneumonia include dementia, seizure disorders, heart failure, cerebrovascular disease, alcoholism, tobacco smoking, chronic obstructive pulmonary disease (COPD), and HIV infection. CA-MRSA pneumonia is more likely in patients with skin colonization or infection with CA-MRSA. Enterobacteriaceae tend to infect patients who have recently been hospitalized and/or received antibiotic therapy or who have comorbidities such as alcoholism, heart failure, or renal failure. *P. aeruginosa* is a particular problem in patients with severe structural lung disease, such as bronchiectasis, cystic fibrosis, or severe COPD. Risk factors for *Legionella* infection include diabetes, hematologic malignancy, cancer, severe renal disease, HIV infection, smoking, male gender, and a recent hotel stay or ship cruise.

CLINICAL MANIFESTATIONS

CAP can vary from indolent to fulminant in presentation and from mild to fatal in severity. Manifestations of progression and severity include both constitutional findings and those limited to the lung and associated structures.

The patient is frequently febrile with tachycardia or may have a history of chills and/or sweats. Cough may be either nonproductive or productive of mucoid, purulent, or blood-tinged sputum. Gross hemoptysis is suggestive of CA-MRSA pneumonia. Depending on severity, the patient may be able to speak in full sentences or may be very short of breath. If the pleura is involved, the patient may experience pleuritic chest pain. Up to

20% of patients may have gastrointestinal symptoms such as nausea, vomiting, and/or diarrhea. Other symptoms may include fatigue, headache, myalgias, and arthralgias.

Findings on physical examination vary with the degree of pulmonary consolidation and the presence or absence of a significant pleural effusion. An increased respiratory rate and use of accessory muscles of respiration are common. Palpation may reveal increased or decreased tactile fremitus, and the percussion note can vary from dull to flat, reflecting underlying consolidated lung and pleural fluid, respectively. Crackles, bronchial breath sounds, and possibly a pleural friction rub may be heard on auscultation. The clinical presentation may not be so obvious in the elderly, who may initially display new-onset or worsening confusion and few other manifestations. Severely ill patients may have septic shock and evidence of organ failure.

The risk of cardiac complications secondary to enhanced inflammation and procoagulant activity is increased. These complications include myocardial infarction, congestive heart failure, and arrhythmias, particularly in the elderly. In pneumococcal CAP, the increased risk of acute coronary events may be partially driven by pneumolysis, which increases platelet activation. Up to 90% of acute coronary syndromes occur in the first week after onset of CAP, and the risk of new-onset congestive heart failure in elderly hospitalized CAP patients can extend up to 1 year.

DIAGNOSIS

When confronted with possible CAP, the physician must ask two questions: Is this pneumonia, and, if so, what is the likely etiology? The former question is typically answered by clinical and radiographic methods, whereas the latter requires the aid of laboratory techniques.

Clinical Diagnosis

The differential diagnosis includes both infectious and noninfectious entities such as acute bronchitis, acute exacerbations of chronic bronchitis, heart failure, pulmonary embolism, hypersensitivity pneumonitis, and radiation pneumonitis. The importance of a careful history cannot be overemphasized. For example, known cardiac disease may suggest worsening pulmonary edema, while underlying carcinoma may suggest lung injury secondary to irradiation.

Unfortunately, the sensitivity and specificity of the findings on physical examination are less than ideal, averaging 58% and 67%, respectively. As mentioned earlier, the elderly may initially present with confusion alone. Therefore, chest radiography is often necessary to differentiate CAP from other conditions. Radiographic findings may include risk factors for increased severity (e.g., cavitation or multilobar involvement). Occasionally, radiographic results suggest an etiologic diagnosis. For example, pneumatoceles suggest infection with *S. aureus*, and an upper-lobe cavitating lesion suggests tuberculosis. CT may be of value in a patient with suspected postobstructive pneumonia caused by a tumor or foreign body or suspected cavitary disease. For outpatients, the clinical and radiologic assessments are usually all that is done before treatment for CAP is started since most laboratory results are not available soon enough to

influence initial management significantly. In certain cases, the availability of rapid point-of-care outpatient diagnostic tests can be very important; for example, rapid diagnosis of influenza virus infection can prompt specific anti-influenza drug treatment and secondary prevention.

Etiologic Diagnosis

The etiology of pneumonia usually cannot be determined solely on the basis of clinical presentation. Except for CAP patients admitted to the ICU, no data exist to show that treatment directed at a specific pathogen is statistically superior to empirical therapy. The benefit of establishing a microbial etiology can therefore be questioned, particularly in light of the cost of diagnostic testing. However, a number of reasons can be advanced for attempting an etiologic diagnosis. Identification of an unexpected pathogen allows narrowing of the initial empirical regimen, thereby decreasing antibiotic selection pressure and lessening the risk of resistance. Pathogens with important public safety implications, such as *Mycobacterium tuberculosis* and influenza virus, may be found in some cases. Finally, without culture and susceptibility data, trends in resistance cannot be followed accurately, and appropriate empirical therapeutic regimens are harder to devise.

Gram's Stain and Culture of Sputum

The main purpose of the sputum Gram's stain is to ensure that a sample is suitable for culture. However, Gram's staining may also identify certain pathogens (e.g., *S. pneumoniae*, *S. aureus*, and gram-negative bacteria) by their characteristic appearance. To be adequate for culture, a sputum sample must have >25 neutrophils and <10 squamous epithelial cells per low-power field. The sensitivity and specificity of the sputum Gram's stain and culture are highly variable. Even in cases of proven bacteremic pneumococcal pneumonia, the yield of positive cultures from sputum samples is ≤50%.

Many patients, particularly elderly individuals, may not be able to produce an appropriate expectorated sputum sample. Others may already have started a course of antibiotics that can interfere with culture results at the time a sample is obtained. Inability to produce sputum can result from dehydration, and its correction may result in increased sputum production and a more obvious infiltrate on chest radiography. For patients admitted to the ICU and intubated, a deep-suction aspirate or bronchoalveolar lavage sample (obtained either via bronchoscopy or non-bronchoscopically) has a high yield on culture when sent to the microbiology laboratory as soon as possible. Since the etiologies in severe CAP are somewhat different from those in milder disease (Table 121-2), the greatest benefit of staining and culturing respiratory secretions is to alert the physician of unsuspected and/or resistant pathogens and to permit appropriate modification of therapy. Other stains and cultures (e.g., specific stains for *M. tuberculosis* or fungi) may be useful as well.

Blood Cultures

The yield from blood cultures, even when samples are collected before antibiotic therapy, is disappointingly low. Only 5–14% of cultures of blood from patients hospitalized with CAP are positive, and the most frequently isolated pathogen is *S. pneumoniae*. Since recommended empirical regimens all provide pneumococcal coverage, a blood culture positive for this pathogen has little, if any, effect on clinical outcome. However, susceptibility data may allow narrowing of antibiotic therapy in appropriate cases. Because of the low yield and the lack of significant impact on outcome, blood cultures are no longer

considered *de rigueur* for all hospitalized CAP patients. Certain high-risk patients—including those with neutropenia secondary to pneumonia, asplenia, complement deficiencies, chronic liver disease, or severe CAP—should have blood cultured.

Urinary Antigen Tests

Two commercially available tests detect pneumococcal and *Legionella* antigen in urine. The test for *Legionella pneumophila* detects only serogroup 1, but this serogroup accounts for most community-acquired cases of Legionnaires' disease in the United States. The sensitivity and specificity of the *Legionella* urine antigen test are as high as 70% and 99%, respectively. The pneumococcal urine antigen test is also quite sensitive and specific (70% and >90%, respectively). Although false-positive results can be obtained with samples from pneumococcus-colonized children, the test is generally reliable. Both tests can detect antigen even after the initiation of appropriate antibiotic therapy.

Polymerase Chain Reaction

PCR tests, which amplify a microorganism's DNA or RNA, are available for a number of pathogens. PCR of nasopharyngeal swabs, for example, has become the standard for diagnosis of respiratory viral infection. In addition, PCR can detect the nucleic acid of *Legionella* species, *M. pneumoniae*, *C. pneumoniae*, and mycobacteria. The cost-effectiveness of PCR testing, however, has not been definitively established. In patients with pneumococcal pneumonia, an increased bacterial load documented in whole blood by PCR is associated with an increased risk of septic shock, the need for mechanical ventilation, and death. Clinical availability of such a test could conceivably help identify patients suitable for ICU admission.

Serology

A fourfold rise in specific IgM antibody titer between acute- and convalescent-phase serum samples is generally considered diagnostic of infection with the pathogen in question. In the past, serologic tests were used to help identify atypical pathogens as well as selected unusual organisms such as *Coxiella burnetii*. Recently, however, they have fallen out of favor because of the time required to obtain a final result for the convalescent-phase sample and the difficulty of interpretation.

Biomarkers

A number of substances can serve as markers of severe inflammation. The two most commonly in use are C-reactive protein (CRP) and procalcitonin (PCT). Levels of these acute-phase reactants increase in the presence of an inflammatory response, particularly to bacterial pathogens. CRP may be of use in the identification of worsening disease or treatment failure, and PCT may play a role in distinguishing bacterial from viral infection, determining the need for antibacterial therapy, or deciding when to discontinue treatment. PCT testing can result in less antibiotic use in CAP with no concomitant increase in treatment failure or mortality risk. These tests should not be used on their own, but, when interpreted in conjunction with other findings from the history, physical examination, radiology, and laboratory tests, may help with antibiotic stewardship and appropriate management of seriously ill patients with CAP.

TREATMENT

TREATMENT

Community-Acquired Pneumonia

SITE OF CARE

The cost of inpatient management exceeds that of outpatient treatment by a factor of 20, and hospitalization accounts for most CAP-related expenditures. Thus the decision to hospitalize a patient with CAP has considerable implications, and late admission to the ICU is associated with increased mortality risk. Certain patients can be managed at home, and others clearly require treatment in the hospital, but the choice is sometimes difficult. Tools that objectively assess the risk of adverse outcomes, including severe illness and death, can minimize unnecessary hospital admissions. Although a number of prediction rules exist, the two most frequently used are the Pneumonia Severity Index (PSI), a prognostic model used to identify patients at low risk of dying, and the CURB-65 criteria, a severity-of-illness score.

To determine the PSI, points are given for 20 variables, including age, coexisting illness, and abnormal physical and laboratory findings. On the basis of the resulting score, patients are assigned to one of five classes with the following mortality rates: class 1, 0.1%; class 2, 0.6%; class 3, 2.8%; class 4, 8.2%; and class 5, 29.2%. Determination of the PSI is often impractical in a busy emergency-department setting because of the number of variables. However, clinical trials demonstrate that routine use of the PSI results in lower admission rates for class 1 and class 2 patients. Patients in class 3 could ideally be admitted to an observation unit until a further decision can be made.

The CURB-65 criteria include five variables: confusion (C); urea >7 mmol/L (U); respiratory rate ≥ 30 /min (R); blood pressure, systolic ≤ 90 mmHg or diastolic ≤ 60 mmHg (B); and age ≥ 65 years. Patients with a score of 0, among whom the 30-day mortality rate is 1.5%, can be treated outside the hospital. With a score of 1 or 2, the patient should be hospitalized unless the score is entirely or in part attributable to an age of ≥ 65 years. In such cases, hospitalization may not be necessary. Among patients with scores of ≥ 3 , mortality rates are 22% overall; these patients may require ICU admission.

It is not clear which assessment tool is superior. Whichever system is used, these objective criteria must always be tempered by careful consideration of factors relevant to individual patients, including the ability to comply reliably with an oral antibiotic regimen and the resources available to the patient outside the hospital.

Neither PSI nor CURB-65 is accurate in determining the need for ICU admission. Septic shock or respiratory failure in the emergency department is an obvious indication for ICU care. However, mortality rates are higher among less ill patients who are admitted to the floor and then deteriorate than among equally ill patients monitored in the ICU. A variety of scores have been proposed to identify patients most likely to have early deterioration ([Table 121-4](#)). Most factors in these scores are similar to the minor severity criteria proposed by the Infectious Diseases Society of America (IDSA) and the American Thoracic Society (ATS) in their guidelines for the management of CAP. Recent data suggest that thrombocytopenia, leukopenia, and hypothermia can be removed from the list of minor criteria.

ANTIBIOTIC RESISTANCE

Antimicrobial resistance is a significant problem that threatens to diminish our therapeutic armamentarium. Misuse of antibiotics results in increased antibiotic selection pressure that can affect resistance locally and globally by clonal dissemination. For CAP, the main resistance issues currently involve *S. pneumoniae* and CA-MRSA.

S. pneumoniae

In general, pneumococcal resistance is acquired by direct DNA incorporation and remodeling resulting from contact with closely related oral commensal bacteria, by the process of natural transformation, or by mutation of certain genes.

The minimal inhibitory concentration (MIC) cutoffs for penicillin in pneumonia are ≤ 2 $\mu\text{g/mL}$ for susceptible, >2 – 4 $\mu\text{g/mL}$ for intermediate, and ≥ 8 $\mu\text{g/mL}$ for resistant. A change in susceptibility thresholds resulted in a dramatic decrease in the proportion of pneumococcal isolates considered nonsusceptible. For meningitis, MIC thresholds remain at the former higher levels. Fortunately, resistance to penicillin appeared to plateau even before the change in MIC thresholds. Pneumococcal resistance to β -lactam drugs is due solely to low-affinity penicillin-binding proteins. Risk factors for penicillin-resistant pneumococcal infection include recent antimicrobial therapy, an age of <2 years or >65 years, attendance at day-care centers, recent hospitalization, and HIV infection.



In contrast to penicillin resistance, resistance to macrolides is increasing through several mechanisms. *Target-site modification* caused by ribosomal methylation in 23S rRNA encoded by the *ermB* gene results in high-level resistance (MICs, ≥ 64 $\mu\text{g/mL}$) to macrolides, lincosamides, and streptogramin B-type antibiotics. The *efflux mechanism* encoded by the *mef* gene (*M phenotype*) is usually associated with low-level resistance (MICs, 1–32 $\mu\text{g/mL}$). These two mechanisms account for $\sim 45\%$ and $\sim 65\%$, respectively, of resistant pneumococcal isolates in the United States. High-level resistance to macrolides is more common in Europe, whereas lower-level resistance predominates in North America. In some countries, including the United States, the prevalence of macrolide-resistant *S. pneumoniae* exceeds 25%. In such situations, a macrolide should not be used as empirical monotherapy.

Pneumococcal resistance to fluoroquinolones (e.g., [ciprofloxacin](#) and [levofloxacin](#)) has been reported. Changes can occur in one or both target sites (topoisomerases II and IV) from mutations in the *gyrA* and *parC* genes, respectively. In addition, an efflux pump may play a role in pneumococcal resistance to fluoroquinolones.

Isolates resistant to drugs from three or more antimicrobial classes with different mechanisms of action are considered MDR strains. The propensity for an association of pneumococcal resistance to penicillin with reduced susceptibility to other drugs, such as macrolides, tetracyclines, and trimethoprim-sulfamethoxazole, is also of concern. In the United States, 58.9% of penicillin-resistant pneumococcal isolates from blood are also resistant to macrolides.

The most important risk factor for antibiotic-resistant pneumococcal infection is use of a specific antibiotic within the previous 3 months. Therefore, a patient's history of prior antibiotic treatment is a critical factor in avoiding the use of an inappropriate antibiotic.

M. pneumoniae

Macrolide-resistant *M. pneumoniae* has been reported in a number of countries, including Germany (3%), Japan (30%), China (95%), and France and the United States (5–13%). *Mycoplasma* resistance to macrolides is on the rise as a result of binding-site mutation in domain V of 23S rRNA.

CA-MRSA

CAP due to MRSA may be caused by the classic hospital-acquired strains or by genotypically and phenotypically distinct community-acquired strains. Most infections with the former strains have been acquired either directly or indirectly by contact with the health care environment (Table 121-1). However, in some hospitals, CA-MRSA strains are displacing the classic hospital-acquired strains—a trend suggesting that the newer strains may be more robust and blurring this distinction.



Methicillin resistance in *S. aureus* is determined by the *mecA* gene, which encodes for resistance to all β -lactam drugs. At least five *staphylococcal chromosomal cassette mec* (*SCCmec*) types have been described. The typical hospital-acquired strain usually has type II or III, whereas CA-MRSA has a type IV *SCCmec* element. CA-MRSA isolates tend to be less resistant than the older hospital-acquired strains and are often susceptible to trimethoprim-sulfamethoxazole, clindamycin, and tetracycline in addition to vancomycin and linezolid. However, the most important distinction is that CA-MRSA strains also carry genes for superantigens, such as enterotoxins B and C and Panton-Valentine leukocidin, a membrane-tropic toxin that can create cytolytic pores in polymorphonuclear neutrophils, monocytes, and macrophages.

Gram-Negative Bacilli

A detailed discussion of resistance among gram-negative bacilli is beyond the scope of this chapter (see Chap. 156). Fluoroquinolone resistance among isolates of *Escherichia coli* from the community appears to be increasing. *Enterobacter* species are typically resistant to cephalosporins; the drugs of choice for use against these bacteria are usually fluoroquinolones or carbapenems. Similarly, when infections due to bacteria producing extended-spectrum β -lactamases are documented or suspected, a fluoroquinolone or a carbapenem should be used.

INITIAL ANTIBIOTIC MANAGEMENT

Since the etiology of CAP is rarely known at the outset of treatment, initial therapy is usually empirical, designed to cover the most likely pathogens (Table 121-2). In all cases, antibiotic treatment should be initiated as expeditiously as possible. The CAP treatment guidelines in the United States (summarized in Table 121-5) represent joint statements from the IDSA and the ATS; the Canadian guidelines come from the Canadian Infectious Disease Society and the Canadian Thoracic Society. In these guidelines, coverage is always provided for the pneumococcus and atypical pathogens. In contrast, guidelines from some European countries do not always include atypical coverage based on local epidemiologic data. The U.S./Canadian

approach is supported by retrospective data derived from administrative databases including thousands of patients. Atypical pathogen coverage provided by the addition of a macrolide to a β -lactam or by the use of a fluoroquinolone alone has been consistently associated with a significant reduction in mortality rates compared with those for β -lactam coverage alone.

For the treatment of severe CAP, accumulating data continue to demonstrate the benefits of including a macrolide, such as reduced mortality. However, two recent studies of patients hospitalized with moderate CAP yielded differing results. One study demonstrated a more rapid return to clinical stability and fewer adverse events with a β -lactam–macrolide combination than with a β -lactam alone. Using cluster randomization, the second study reported no difference among three regimens—a β -lactam alone, a β -lactam–macrolide combination, and a fluoroquinolone—but had significant design flaws, including a lack of chest radiographic confirmation in 24% of cases and significant rates of noncompliance with the assigned regimen.

Empirical treatment regimens for CAP are listed in [Table 121-5](#). In general, the recommendations in the IDSA/ATS guidelines published in 2007 continue to apply but with a possible exception for treatment of outpatients who have previously been well and have not received an antibiotic within 3 months. Given the rise of macrolide resistance among pneumococci, consideration of local epidemiologic and susceptibility data as well as the patient's recent use of any antibiotics is imperative before selection of a regimen, particularly as regards macrolide monotherapy. If concern exists about macrolide resistance, the patient is otherwise well and has not recently received antibiotics, and the local doxycycline resistance rate among pneumococcal isolates is <25%, doxycycline may be used instead of macrolide monotherapy. Otherwise, a fluoroquinolone or a β -lactam plus a macrolide should be used.

A meta-analysis found ceftaroline to be superior to ceftriaxone as the β -lactam component of IV empirical treatment of CAP in hospitalized patients in PORT risk class III or IV who have not received prior antibiotics. Clinical response rates for patients infected with *S. pneumoniae* or *S. aureus* also favored ceftaroline. Patients who had documented or suspected infection due to *P. aeruginosa* were excluded.

Once the etiologic agent(s) and their susceptibilities are known, therapy may be altered to target the specific pathogen(s). However, this decision is not always straightforward. If blood cultures yield *S. pneumoniae* sensitive to penicillin after 2 days of treatment with a macrolide plus a β -lactam or with a fluoroquinolone alone, should therapy be switched to penicillin alone? The concern here is that a β -lactam alone would not be effective in the potential 15% of cases with atypical co-infection. No standard approach exists. Some experts think that 3 days of macrolide therapy is adequate for *Mycoplasma* infection and that, unless the test for *Legionella* urinary antigen is positive, treatment can be continued with a β -lactam alone. In all cases, the individual patient and the various risk factors must be considered.

Management of bacteremic pneumococcal pneumonia is also controversial. Data from nonrandomized studies suggest that combination therapy (especially with a β -lactam–macrolide combination) is associated with a lower mortality rate than monotherapy, particularly in severely ill patients. The exact reason is

unknown, but possible explanations include an additive or synergistic antibacterial effect, antimicrobial tolerance, atypical co-infection, or the immunomodulatory effects of the macrolides.

For CAP patients admitted to the ICU, the risk of infection with *P. aeruginosa* or CA-MRSA is increased. Empirical coverage should be considered when a patient has risk factors or a Gram's stain suggestive of these pathogens (Table 121-5). If CA-MRSA is suspected, either linezolid or vancomycin—with or without clindamycin to inhibit toxin production—can be added to the initial empirical regimen. There is increasing concern about vancomycin's loss of potency against MRSA, poor penetration into epithelial lining fluid, and lack of effect on toxin production relative to linezolid.

Although hospitalized patients have traditionally received initial therapy by the IV route, some drugs—particularly the fluoroquinolones—are very well absorbed and can be given orally from the outset to certain patients. For patients initially treated IV, a switch to oral treatment is appropriate as long as the patient can ingest and absorb the drugs, is hemodynamically stable, and is showing clinical improvement.

The duration of treatment for CAP has generated considerable interest. Studies with fluoroquinolones and telithromycin suggest that a 5-day course is sufficient for otherwise uncomplicated CAP but a longer course may be required for patients with bacteremia, metastatic infection, or infection with a virulent pathogen such as *P. aeruginosa* or CA-MRSA.

ADJUNCTIVE MEASURES

In addition to appropriate antimicrobial therapy, certain adjunctive measures should be used. Adequate hydration, oxygen therapy for hypoxemia, vasopressors, and assisted ventilation when necessary are critical to successful treatment. Randomized placebo-controlled trials have shown a benefit in treatment of hospitalized patients and patients who have severe CAP with prednisone and methylprednisolone, respectively. The value of adjunctive therapy with agents such as statins and angiotensin-converting enzyme inhibitors remains unproven in the management of CAP.

FAILURE TO IMPROVE

Patients slow to respond to therapy should be reevaluated at about day 3 (sooner if their condition is worsening rather than simply not improving), and several possible scenarios should be considered. A number of noninfectious conditions mimic pneumonia, including pulmonary edema, pulmonary embolism, lung carcinoma, radiation and hypersensitivity pneumonitis, and connective tissue disease involving the lungs. If the patient truly has CAP and empirical treatment is aimed at the correct pathogen, lack of response may be explained in a number of ways. The pathogen may be resistant to the drug selected, or a sequestered focus (e.g., lung abscess or empyema) may be blocking access of the antibiotic(s) to the pathogen. The patient may be getting either the wrong drug or the correct drug at the wrong dose or frequency of administration. Another possibility is that CAP is the correct diagnosis but an unsuspected pathogen (e.g., CA-MRSA, *M. tuberculosis*, or a fungus) is the cause. Nosocomial superinfections—both pulmonary and extrapulmonary—are other possible explanations for a hospitalized patient's failure to improve or deterioration. In all cases of delayed response or worsening condition, the patient must be carefully reassessed and appropriate studies initiated, possibly including procedures such as CT or bronchoscopy.

COMPLICATIONS

Complications of severe CAP include respiratory failure, shock and multiorgan failure, coagulopathy, and exacerbation of comorbid illnesses. Three particularly noteworthy conditions are metastatic infection, lung abscess, and complicated pleural effusion. Metastatic infection (e.g., brain abscess or endocarditis) is very unusual and will require a high degree of suspicion and a detailed workup for proper treatment. Lung abscess may occur in association with aspiration or with infection caused by a single CAP pathogen, such as CA-MRSA, *P. aeruginosa*, or (rarely) *S. pneumoniae*. Aspiration pneumonia is typically a polymicrobial infection involving both aerobes and anaerobes. A significant pleural effusion should be tapped for both diagnostic and therapeutic purposes. If the fluid has a pH of <7, a glucose level of <2.2 mmol/L, and a lactate dehydrogenase concentration of >1000 U/L or if bacteria are seen or cultured, it should be completely drained; a chest tube is often required, and video-assisted thoracoscopy may be needed for late treatment or difficult cases.

FOLLOW-UP

Fever and leukocytosis usually resolve within 2–4 days in otherwise healthy patients with CAP, but physical findings may persist longer. Chest radiographic abnormalities are slowest to resolve (4–12 weeks), with the speed of clearance depending on the patient's age and underlying lung disease. Patients may be discharged from the hospital once their clinical conditions, including comorbidities, are stable. The site of residence after discharge (nursing home, home with family, home alone) is an important discharge consideration, particularly for elderly patients. For a hospitalized patient, a follow-up radiograph ~4–6 weeks later is recommended. If relapse or recurrence is documented, particularly in the same lung segment, the possibility of an underlying neoplasm must be considered.

TABLE 121-4

Risk Factors for Early Deterioration in Community-Acquired Pneumonia

Multilobar infiltrates	Hypoalbuminemia
Severe hypoxemia (arterial saturation <90%)	Neutropenia
Severe acidosis (pH <7.30)	Thrombocytopenia
Mental confusion	Hyponatremia
Severe tachypnea (>30 breaths/min)	Hypoglycemia

TABLE 121-5

Empirical Antibiotic Treatment of Community-Acquired Pneumonia

Outpatients
<ol style="list-style-type: none"> Previously healthy and no antibiotics in past 3 months <ul style="list-style-type: none"> A macrolide [clarithromycin (500 mg PO bid) or azithromycin (500 mg PO once, then 250 mg qd)] <i>or</i> Doxycycline (100 mg PO bid) Comorbidities or antibiotics in past 3 months: select an alternative from a different class <ul style="list-style-type: none"> A respiratory fluoroquinolone [moxifloxacin (400 mg PO qd), gemifloxacin (320 mg PO qd), levofloxacin (750 mg PO qd)] <i>or</i> A β-lactam [preferred: high-dose amoxicillin (1 g tid) or amoxicillin/clavulanate (2 g bid); alternatives: ceftriaxone (1–2 g IV qd), cefpodoxime (200 mg PO bid), or cefuroxime (500 mg PO bid)] <i>plus</i> a macrolide^a In regions with a high rate of “high-level” pneumococcal macrolide resistance,^b consider alternatives listed above for patients with comorbidities.
Inpatients, Non-ICU
<ul style="list-style-type: none"> A respiratory fluoroquinolone [e.g., moxifloxacin (400 mg PO or IV qd) or levofloxacin (750 mg PO or IV qd)] A β-lactam^c [e.g., ceftriaxone (1–2 g IV qd), ampicillin (1–2 g IV q4–6h), cefotaxime (1–2 g IV q8h), ertapenem (1 g IV qd)] <i>plus</i> a macrolide^d [e.g., oral clarithromycin or azithromycin as listed above or IV azithromycin (1 g once, then 500 mg qd)]
Inpatients, ICU
<ul style="list-style-type: none"> A β-lactam^e [e.g., ceftriaxone (2 g IV qd), ampicillin-sulbactam (2 g IV q8h), or cefotaxime (1–2 g IV q8h)] <i>plus</i> either azithromycin or a fluoroquinolone (as listed above for inpatients, non-ICU)
Special Concerns
<p><i>If Pseudomonas is a consideration:</i></p> <ul style="list-style-type: none"> An antipseudomonal β-lactam [e.g., piperacillin/tazobactam (4.5 g IV q6h), cefepime (1–2 g IV q12h), imipenem (500 mg IV q6h), meropenem (1 g IV q8h)] <i>plus</i> either ciprofloxacin (400 mg IV q12h) or levofloxacin (750 mg IV qd) The above β-lactams <i>plus</i> an aminoglycoside [amikacin (15 mg/kg qd) or tobramycin (1.7 mg/kg qd)] <i>plus</i> azithromycin The above β-lactams^f <i>plus</i> an aminoglycoside <i>plus</i> an antipneumococcal fluoroquinolone

If CA-MRSA is a consideration:

- Add linezolid (600 mg IV q12h) or vancomycin (15 mg/kg q12h initially, with adjusted doses) *plus* [clindamycin](#) (300 mg q6h)

a

Doxycycline (100 mg PO bid) is an alternative to the macrolide.

b

MICs >16 µg/mL in 25% of isolates.

c

A respiratory fluoroquinolone should be used for penicillin-allergic patients.

d

Doxycycline (100 mg IV q12h) is an alternative to the macrolide.

e

For penicillin-allergic patients, use a respiratory fluoroquinolone and [aztreonam](#) (2 g IV q8h).

f

For penicillin-allergic patients, substitute [aztreonam](#).

Abbreviations: CA-MRSA, community-acquired methicillin-resistant *Staphylococcus aureus*; ICU, intensive care unit.

PROGNOSIS

The prognosis of CAP depends on the patient's age, comorbidities, and site of treatment (inpatient or outpatient). Young patients without comorbidity do well and usually recover fully after ~2 weeks. Older patients and those with comorbid conditions can take several weeks longer to recover fully. The overall mortality rate for the outpatient group is <5%. For patients requiring hospitalization, the overall mortality rate ranges from 2% to 40%, depending on the category of patient and the processes of care, particularly the administration of appropriate antibiotics as soon as possible.

PREVENTION

The main preventive measure is vaccination ([Chap. 118](#)). Recommendations of the Advisory Committee on Immunization Practices should be followed for influenza and pneumococcal vaccines.

A pneumococcal polysaccharide vaccine (PPSV23) and a protein conjugate pneumococcal vaccine (PCV13) are available in the United States ([Chap. 141](#)). The former product contains capsular material from 23 pneumococcal serotypes; in the latter, capsular polysaccharide from 13 of the most common pneumococcal pathogens affecting children is linked to an immunogenic protein. PCV13 produces T cell-dependent antigens that result in long-term immunologic memory. Administration of this vaccine to children has led to an overall decrease in the prevalence of antimicrobial-resistant pneumococci and in the incidence of invasive

pneumococcal disease among both children and adults. However, vaccination can be followed by the replacement of vaccine serotypes with nonvaccine serotypes, as was seen with serotypes 19A and 35B after introduction of the original 7-valent conjugate vaccine. PCV13 is also recommended for the elderly and for younger immunocompromised patients. Because of an increased risk of pneumococcal infection, even among patients without obstructive lung disease, smokers should be strongly encouraged to stop smoking.

The influenza vaccine is available in an inactivated or recombinant form. The live attenuated influenza vaccine or “nasal spray” vaccine is no longer recommended. In the event of an influenza outbreak, unprotected patients at risk from complications should be vaccinated immediately and given chemoprophylaxis with either oseltamivir or zanamivir for 2 weeks—i.e., until vaccine-induced antibody levels are sufficiently high.

VENTILATOR-ASSOCIATED PNEUMONIA

Most research on hospital-acquired pneumonia has focused on VAP. However, the information and principles based on this research can be applied to non-ICU HAP as well. The greatest difference between VAP and HAP studies is the dependence on expectorated sputum for a microbiologic diagnosis of HAP (as for that of CAP), which is further complicated by frequent colonization by pathogens in patients with HAP. Therefore, most of the literature has focused on HAP resulting in intubation, where, once again, access to the lower respiratory tract facilitates an etiologic diagnosis.

ETIOLOGY

Potential etiologic agents of VAP include both MDR and non-MDR bacterial pathogens ([Table 121-6](#)). The non-MDR group is nearly identical to the pathogens found in severe CAP ([Table 121-2](#)); it is not surprising that such pathogens predominate if VAP develops in the first 5–7 days of the hospital stay. However, if patients have other risk factors, MDR pathogens are a consideration, even early in the hospital course. The relative frequency of individual MDR pathogens can vary significantly from hospital to hospital and even between different critical care units within the same institution. Most hospitals have problems with *P. aeruginosa* and MRSA, but other MDR pathogens are often institution-specific. Less commonly, fungal and viral pathogens cause VAP, usually affecting severely immunocompromised patients. Rarely, community-associated viruses cause mini-epidemics, usually when introduced by ill health care workers.

TABLE 121-6

Microbiologic Causes of Ventilator-Associated Pneumonia

NON-MDR PATHOGENS	MDR PATHOGENS
<i>Streptococcus pneumoniae</i> Other <i>Streptococcus</i> spp. <i>Haemophilus influenzae</i> Methicillin-sensitive <i>Staphylococcus aureus</i> Antibiotic-sensitive Enterobacteriaceae <i>Escherichia coli</i> <i>Klebsiella pneumoniae</i> <i>Proteus</i> spp. <i>Enterobacter</i> spp. <i>Serratia marcescens</i>	<i>Pseudomonas aeruginosa</i> Methicillin-resistant <i>S. aureus</i> <i>Acinetobacter</i> spp. Antibiotic-resistant Enterobacteriaceae ESBL-positive strains Carbapenem-resistant strains <i>Legionella pneumophila</i> <i>Burkholderia cepacia</i> <i>Aspergillus</i> spp.

Abbreviations: ESBL, extended-spectrum β -lactamase; MDR, multidrug-resistant.

EPIDEMIOLOGY

Pneumonia is a common complication among patients requiring mechanical ventilation. Prevalence estimates vary between 6 and 52 cases per 100 patients, depending on the population studied. On any given day in the ICU, an average of 10% of patients will have pneumonia—VAP in the overwhelming majority of cases. The frequency of diagnosis is not static but changes with the duration of mechanical ventilation, with the highest hazard ratio in the first 5 days and a plateau in additional cases (1% per day) after ~2 weeks. However, the cumulative rate among patients who remain ventilated for as long as 30 days is as high as 70%. These rates often do not reflect the recurrence of VAP in the same patient. Once a ventilated patient is transferred to a chronic-care facility or to home, the incidence of pneumonia drops significantly, especially in the absence of other risk factors for pneumonia. However, in chronic ventilator units, purulent tracheobronchitis becomes a significant issue, often interfering with efforts to wean patients off mechanical ventilation ([Chap. 295](#)).

Three factors are critical in the pathogenesis of VAP: colonization of the oropharynx with pathogenic microorganisms, aspiration of these organisms from the oropharynx into the lower respiratory tract, and compromise of the normal host defense mechanisms. Most risk factors and their corresponding prevention strategies pertain to one of these three factors ([Table 121-7](#)).

TABLE 121-7

Pathogenic Mechanisms and Corresponding Prevention Strategies for Ventilator-Associated Pneumonia

PATHOGENIC MECHANISM	PREVENTION STRATEGY
Oropharyngeal colonization with pathogenic bacteria	
Elimination of normal flora	Avoidance of prolonged antibiotic courses
Large-volume oropharyngeal aspiration around time of intubation	Short course of prophylactic antibiotics for comatose patients ^a
Gastroesophageal reflux	Postpyloric enteral feeding ^b ; avoidance of high gastric residuals, prokinetic agents
Bacterial overgrowth of stomach	Avoidance of prophylactic agents that raise gastric pH ^b ; selective decontamination of digestive tract with nonabsorbable antibiotics ^b
Cross-infection from other colonized patients	Hand washing, especially with alcohol-based hand rub; intensive infection control education ^a ; isolation; proper cleaning of reusable equipment
Large-volume aspiration	Endotracheal intubation; rapid-sequence intubation technique; avoidance of sedation; decompression of small-bowel obstruction
Microaspiration around endotracheal tube	
Endotracheal intubation	Noninvasive ventilation ^a
Prolonged duration of ventilation	Daily awakening from sedation, ^a weaning protocols ^a
Abnormal swallowing function	Early percutaneous tracheostomy ^a

PATHOGENIC MECHANISM	PREVENTION STRATEGY
Secretions pooled above endotracheal tube	Head of bed elevated ^a ; continuous aspiration of subglottic secretions with specialized endotracheal tube ^a ; avoidance of reintubation; minimization of sedation and patient transport
Altered lower respiratory host defenses	Tight glycemic control ^b ; lowering of hemoglobin transfusion threshold

a

Strategies demonstrated to be effective in at least one randomized controlled trial.

b

Strategies with negative randomized trials or conflicting results.

The most obvious risk factor is the endotracheal tube, which bypasses the normal mechanical factors preventing aspiration. While the presence of an endotracheal tube may prevent large-volume aspiration, microaspiration is actually exacerbated by secretions pooling above the cuff. The endotracheal tube and the concomitant need for suctioning can damage the tracheal mucosa, thereby facilitating tracheal colonization. In addition, pathogenic bacteria can form a glycocalyx biofilm on the tube's surface that protects them from both antibiotics and host defenses. The bacteria can also be dislodged during suctioning and can reinoculate the trachea, or tiny fragments of glycocalyx can embolize to distal airways, carrying bacteria with them.

In a high percentage of critically ill patients, the normal oropharyngeal flora is replaced by pathogenic microorganisms. The most important risk factors are antibiotic selection pressure, cross-infection from other infected/colonized patients or contaminated equipment, and malnutrition. Of these factors, antibiotic exposure poses the greatest risk by far. Pathogens such as *P. aeruginosa* almost never cause infection in patients without prior exposure to antibiotics. The recent emphasis on hand hygiene has lowered the cross-infection rate.

How the lower respiratory tract defenses become overwhelmed remains poorly understood. Almost all intubated patients experience microaspiration and are at least transiently colonized with pathogenic bacteria. However, only around one-third of colonized patients develop VAP. Colony counts increase to high levels, sometimes days before the development of clinical pneumonia; these increases suggest that the final step in VAP development, independent of aspiration and oropharyngeal colonization, is the overwhelming of host defenses. Severely ill patients with sepsis and trauma appear to enter a state of immunoparalysis several days after admission to the ICU—a time that corresponds to the greatest risk of developing VAP. The mechanism of this immunosuppression is not clear, although several factors have been suggested. Hyperglycemia and more frequent transfusions adversely affect the immune response.

CLINICAL MANIFESTATIONS

The clinical manifestations are generally the same in VAP as in all other forms of pneumonia: fever, leukocytosis, increase in respiratory secretions, and pulmonary consolidation on physical examination, along with a new or changing radiographic infiltrate. The frequency of abnormal chest radiographs before the onset of pneumonia in intubated patients and the limitations of portable radiographic technique make interpretation of radiographs more difficult than in patients who are not intubated. Other clinical features may include tachypnea, tachycardia, worsening oxygenation, and increased minute ventilation.

DIAGNOSIS

No single set of criteria is reliably diagnostic of pneumonia in a ventilated patient. The inability to accurately identify such patients compromises efforts to prevent and treat VAP and even calls into question estimates of the impact of VAP on mortality rates.

Application of the clinical criteria typical for CAP consistently results in overdiagnosis of VAP, largely because of three common findings in at-risk patients: (1) frequent tracheal colonization with pathogenic bacteria in patients with endotracheal tubes, (2) multiple alternative causes of radiographic infiltrates in mechanically ventilated patients, and (3) the high frequency of other sources of fever in critically ill patients. The differential diagnosis of VAP includes a number of entities such as atypical pulmonary edema, pulmonary contusion, alveolar hemorrhage, hypersensitivity pneumonitis, acute respiratory distress syndrome, and pulmonary embolism. Clinical findings in ventilated patients with fever and/or leukocytosis may have alternative causes, including antibiotic-associated diarrhea, central line-associated infection, sinusitis, urinary tract infection, pancreatitis, and drug fever. Conditions mimicking pneumonia are often documented in patients in whom VAP has been ruled out by accurate diagnostic techniques. Most of these alternative diagnoses do not require antibiotic treatment; require antibiotics different from those used to treat VAP; or require some additional intervention, such as surgical drainage or catheter removal, for optimal management.

This diagnostic dilemma has led to debate and controversy. The major question is whether a quantitative-culture approach as a means of eliminating false-positive clinical diagnoses is superior to the clinical approach enhanced by principles learned from quantitative-culture studies. The most recent IDSA/ATS guidelines for HAP/VAP gave a weak recommendation for the clinical approach based on availability of resources, cost, and availability of expertise. The guidelines did acknowledge that the use of a quantitative approach may result in less antibiotic use, which may be critical for antibiotic stewardship in the ICU. Therefore, the approach at each institution, or potentially for each patient, should balance the frequency of complex illnesses that are associated with (1) greater frequency of alternative causes of the clinical manifestations, (2) higher colonization rates, and (3) more frequent prior antibiotic therapy versus availability and expertise of invasive techniques with quantitative cultures.

Quantitative-Culture Approach

The essence of the quantitative-culture approach is discrimination between colonization and true infection through determination of the bacterial burden. The more distal in the respiratory tree the diagnostic

sampling, the more specific the results and therefore the lower the threshold of growth necessary to diagnose pneumonia and exclude colonization. For example, a quantitative endotracheal aspirate yields proximate samples, and the diagnostic threshold is 10^6 cfu/mL. The protected specimen brush method, in contrast, obtains distal samples and has a threshold of 10^3 cfu/mL. Conversely, sensitivity declines as more distal secretions are obtained, especially when they are collected blindly (i.e., by a technique other than bronchoscopy). Additional tests that may increase the diagnostic yield include Gram's staining, differential cell counts, staining for intracellular organisms, and detection of local protein levels elevated in response to infection.

The key piece of a quantitative-culture approach is to base subsequent antibiotic therapy on the results of the quantitative cultures. In a study comparing the quantitative with the clinical approach, the use of bronchoscopic quantitative cultures resulted in significantly less antibiotic use at 14 days after study entry, a lower 14-day mortality rate, and a lower 28-day severity-adjusted mortality rate. In addition, more alternative sites of infection were found in patients randomized to the quantitative-culture strategy. A critical aspect of this study was that antibiotic treatment was initiated only in patients whose gram-stained respiratory sample was positive or who displayed signs of hemodynamic instability. Fewer than half as many patients were treated for pneumonia in the bronchoscopy group, and only one-third as many microorganisms were cultured. Other randomized trials of the quantitative-culture approach did not closely link antibiotic management with the results of cultures; thus the validity of their results was compromised.

The Achilles heel of the quantitative approach is the effect of antibiotic therapy. With sensitive microorganisms, a single antibiotic dose can reduce colony counts below the diagnostic threshold. Recent changes in antibiotic therapy are the most significant. After 3 days, the operating characteristics of the tests improve to the point at which they are equivalent to results when no prior antibiotic therapy has been given. Conversely, colony counts above the diagnostic threshold during antibiotic therapy suggest that the current antibiotics are ineffective. Even the normal host response may be sufficient to reduce quantitative-culture counts below the diagnostic threshold if sampling is delayed. In short, expertise in quantitative-culture techniques is critical, with a specimen obtained as soon as pneumonia is suspected and before antibiotic therapy is initiated or changed.

Clinical Approach

General knowledge of the lack of specificity of a clinical diagnosis of VAP and results from invasive quantitative-culture studies have actually improved the clinical approach to the diagnosis of VAP. Tracheal aspirates generally yield at least twice as many potential pathogens as quantitative cultures. However, the causative pathogen is almost always present. The absence of bacteria in gram-stained endotracheal aspirates makes pneumonia an unlikely cause of fever or pulmonary infiltrates. These findings, coupled with a heightened awareness of the alternative diagnoses possible in patients with suspected VAP, can prevent inappropriate overtreatment for pneumonia. Furthermore, the absence of an MDR pathogen in tracheal aspirate cultures eliminates the need for MDR coverage, allowing empirical antibiotic therapy to be de-escalated. Since the main benefits of bronchoscopic quantitative cultures are decreased antibiotic selection

pressure (which reduces the risk of subsequent infection with MDR pathogens) and the identification of alternative sources of infection, a clinical diagnostic approach that incorporates such principles may result in similar outcomes.

TREATMENT

TREATMENT

Ventilator-Associated Pneumonia

Many studies have demonstrated higher mortality rates with initially inappropriate empirical antibiotic therapy. The key to appropriate antibiotic management of VAP is an appreciation of the resistance patterns of the most likely pathogens in a given patient.

ANTIBIOTIC RESISTANCE

If not for the higher risk of infection with MDR pathogens (Table 121-6), VAP could be treated with the same antibiotics used for severe CAP. However, antibiotic selection pressure leads to the frequent involvement of MDR pathogens by selecting either for drug-resistant isolates of common pathogens (MRSA and Enterobacteriaceae producing extended-spectrum β -lactamases or carbapenemases) or for intrinsically resistant pathogens (*P. aeruginosa* and *Acinetobacter* species). Frequent use of β -lactam drugs, especially cephalosporins, appears to be the major risk factor for infection with MRSA and extended-spectrum β -lactamase-positive strains.

P. aeruginosa has demonstrated the ability to develop resistance to all routinely used antibiotics. Unfortunately, even if initially sensitive, *P. aeruginosa* isolates also have a propensity to develop resistance during treatment. Either de-repression of resistance genes or selection of resistant clones within the large bacterial inoculum associated with most pneumonias may be the cause. *Acinetobacter* species, *Stenotrophomonas maltophilia*, and *Burkholderia cepacia* are intrinsically resistant to many of the empirical antibiotic regimens employed (see below). VAP caused by these pathogens emerges during treatment of other infections, and resistance is always evident at initial diagnosis.

EMPIRICAL THERAPY

Recommended options for empirical therapy are listed in Table 121-8. Treatment should be started once diagnostic specimens have been obtained. The major factor in the selection of agents is the presence of risk factors for MDR pathogens. Choices among the various options listed depend on local patterns of resistance and—a very important factor—the patient's prior antibiotic exposure. Knowledge of the local hospital's—and even the specific ICU's—antibiogram and the local incidence of specific MDR pathogens (e.g., MRSA) is critical in selecting appropriate empirical therapy.

The majority of patients *without* risk factors for MDR infection can be treated with a single agent. Unfortunately, the proportion of patients with no MDR risk factors is <10% in some ICUs and is unknown for HAP patients. The major difference from CAP is the markedly lower incidence of atypical pathogens in VAP; the exception is *Legionella*, which can be a nosocomial pathogen, especially with breakdowns in the

treatment of potable water in the hospital. The standard recommendation for patients *with* risk factors for MDR infection is for three antibiotics: two directed at *P. aeruginosa* and one at MRSA. A β -lactam agent provides the greatest coverage, yet even the broadest-spectrum agent—a carbapenem—still provides inappropriate initial therapy in up to 10–15% of cases at some centers. The emergence of carbapenem resistance at some institutions requires the addition of polymyxins to the combination-therapy options.

SPECIFIC TREATMENT

Once an etiologic diagnosis is made, broad-spectrum empirical therapy can be modified to specifically address the known pathogen. For patients with MDR risk factors, antibiotic regimens can be reduced to a single agent in most cases. Only a minority of cases require a complete course with two or three drugs. A negative tracheal-aspirate culture or growth below the threshold for quantitative cultures of samples obtained before any antibiotic change strongly suggests that antibiotics should be discontinued or that a search for an alternative diagnosis should be pursued. Identification of other confirmed or suspected sites of infection may require ongoing antibiotic therapy, but the spectrum of pathogens (and the corresponding antibiotic choices) may be different from those for VAP. A 7- or 8-day course of therapy is just as effective as a 2-week course and is associated with less frequent emergence of antibiotic-resistant strains.

The major controversy regarding specific therapy for VAP concerns the need for ongoing combination treatment of *Pseudomonas* pneumonia. No randomized controlled trials have demonstrated a benefit of combination therapy with a β -lactam and an aminoglycoside, nor have subgroup analyses in other trials found a survival benefit with such a regimen. The unacceptably high rates of clinical failure and death for VAP caused by *P. aeruginosa* despite combination therapy (see “Failure to Improve,” below) indicate that better regimens are needed, perhaps including aerosolized antibiotics. Current guidelines recommend against continued combination therapy for most cases of *Pseudomonas* pneumonia.

FAILURE TO IMPROVE

Treatment failure is not uncommon in VAP, especially that caused by MDR pathogens. VAP caused by MRSA is associated with a 40% clinical failure rate when treated with standard-dose vancomycin. One proposed but unproven solution is the use of high-dose individualized treatment, although the risk of renal toxicity increases with this strategy. In addition, the MIC of vancomycin has been increasing, and a high percentage of clinical failures occur when the MIC is in the upper range of sensitivity (i.e., 1.5–2 $\mu\text{g/mL}$). Linezolid appears to be 15% more efficacious than even adjusted-dose vancomycin and is clearly preferred in patients with renal insufficiency and those infected with high-MIC isolates of MRSA. VAP due to *Pseudomonas* has a 40–50% failure rate, no matter what the regimen. Causes of clinical failure vary with the pathogen(s) and the antibiotic(s). Inappropriate initial therapy can usually be minimized by use of the recommended combination regimen (Table 121-8). However, the emergence of β -lactam resistance during therapy is an important problem, especially in infection with *Pseudomonas* and *Enterobacter* species. Recurrent VAP caused by the same pathogen is possible because the biofilm on endotracheal tubes allows reintroduction of the microorganism. Studies of VAP caused by *Pseudomonas* show that approximately half of recurrent cases are caused by a new strain.

Treatment failure is very difficult to diagnose early in the therapeutic course, and discrimination among the various potential causes is a challenge. Pneumonia due to a new superinfection, the presence of extrapulmonary infection, and drug toxicity must be considered in the differential diagnosis of treatment failure. Serial measurements of procalcitonin levels appear to track the clinical response accurately, while repeat quantitative cultures may clarify the microbiologic response.

COMPLICATIONS

Apart from death, the major complication of VAP is prolongation of mechanical ventilation, with corresponding increases in the duration of ICU stay and hospitalization. In most studies, an additional week of mechanical ventilation resulting from VAP is common. The additional expense of this complication often warrants costly and aggressive efforts at prevention.

In rare cases, necrotizing pneumonia (e.g., that due to *P. aeruginosa*) can cause significant pulmonary hemorrhage. More commonly, necrotizing infections result in the long-term complications of bronchiectasis and parenchymal scarring leading to recurrent pneumonia. Other long-term complications of pneumonia are underappreciated. Pneumonia results in a catabolic state in a patient already nutritionally at risk. The muscle loss and general debilitation from an episode of VAP often require prolonged rehabilitation and, in the elderly, often result in an inability to return to independent function and the need for nursing home placement.

FOLLOW-UP

Clinical improvement, if it occurs, is usually evident within 48–72 h of the initiation of antimicrobial treatment. Because findings on chest radiography often worsen initially during treatment, they are less helpful than clinical criteria as an indicator of clinical response in severe pneumonia.

TABLE 121-8

Empirical Antibiotic Treatment of Hospital-Acquired and Ventilator-Associated Pneumonia

NO RISK FACTORS FOR RESISTANT GRAM-NEGATIVE PATHOGEN	RISK FACTORS FOR RESISTANT GRAM-NEGATIVE PATHOGEN ^a (<i>CHOOSE ONE FROM EACH COLUMN</i>)	
Piperacillin-tazobactam (4.5 g IV q6h ^b) Cefepime (2 g IV q8h) Levofloxacin (750 mg IV q24h)	Piperacillin-tazobactam (4.5 g IV q6h ^b) Cefepime (2 g IV q8h) Ceftazidime (2 g IV q8h) Imipenem (500 mg IV q6h ^b) Meropenem (1 g IV q8h)	Amikacin (15–20 mg/kg IV q24h) Gentamicin (5–7 mg/kg IV q24h) Tobramycin (5–7 mg/kg IV q24h) Ciprofloxacin (400 mg IV q8h) Levofloxacin (750 mg IV q24h) Colistin (loading dose of 5 mg/kg IV followed by maintenance doses of 2.5 mg × [1.5 × CrCl + 30] IV q12h) Polymyxin B (2.5–3.0 mg/kg per day IV in 2 divided doses)
Risk Factors for MRSA^b (<i>Add to above</i>)		
Linezolid (600 mg IV q12h) <i>or</i> Adjusted-dose vancomycin (trough level, 15–20 mg/dL)		

a

Prior antibiotic therapy, prior hospitalization, local antibiogram.

b

Prior antibiotic therapy, prior hospitalization, known MRSA colonization, chronic hemodialysis, local documented MRSA pneumonia rate >10% (or local rate unknown).

Abbreviation: CrCl, creatinine clearance rate; MRSA, methicillin-resistant *Staphylococcus aureus*.**PROGNOSIS**

VAP is associated with crude mortality rates as high as 50–70%, but the real issue is attributable mortality. Many patients with VAP have underlying diseases that would result in death even if VAP did not occur. Attributable mortality exceeded 25% in one matched-cohort study, while more recent studies have suggested much lower rates. Some variability in VAP mortality rates is clearly related to the type of patient and ICU studied. VAP in trauma patients is not associated with attributable mortality, possibly because many of the patients were otherwise healthy before being injured. The causative pathogen also plays a major role. Generally, MDR pathogens are associated with significantly greater attributable mortality than non-MDR pathogens. Pneumonia caused by some pathogens (e.g., *S. maltophilia*) is simply a marker for a patient whose immune system is so compromised that death is almost inevitable.

PREVENTION

Because of the significance of endotracheal intubation as a risk factor for VAP, the most important preventive intervention is to avoid intubation or minimize its duration ([Table 121-7](#)). Successful noninvasive ventilation avoids many of the problems associated with endotracheal tubes. Strategies that minimize the duration of ventilation through daily holding of sedation and formal weaning protocols have also been highly effective in preventing VAP.

Unfortunately, a tradeoff in risks is sometimes necessary. Aggressive attempts to extubate early may result in reintubation(s) and increase aspiration, posing a risk of VAP. Heavy continuous sedation increases VAP risk, but self-extubation because of insufficient sedation is also a risk. The tradeoffs also apply to antibiotic therapy. Short-course antibiotic prophylaxis can decrease the risk of VAP in comatose patients requiring intubation, and data suggest that antibiotics decrease VAP rates overall. However, the major benefit appears to be a decrease in the incidence of early-onset VAP, which is usually caused by the less pathogenic non-MDR microorganisms. Conversely, prolonged courses of antibiotics consistently increase the risk of VAP caused by more lethal MDR pathogens. Despite its virulence and associated mortality, VAP caused by *Pseudomonas* is rare among patients who have not recently received antibiotics.

Minimizing microaspiration around the endotracheal tube cuff is also a strategy for avoidance of VAP. Simply elevating the head of the bed (at least 30° above horizontal but preferably 45°) decreases VAP rates. Specially modified endotracheal tubes that allow removal of the secretions pooled above the cuff may also prevent VAP. The risk-to-benefit ratio of transporting the patient outside the ICU for diagnostic tests or procedures should be carefully considered, since VAP rates are increased among transported patients.

The role played by overgrowth of the normal bowel flora in the stomach in the pathogenesis of VAP is questionable. MRSA and the nonfermenters *P. aeruginosa* and *Acinetobacter* species are not normally part of the bowel flora but reside primarily in the nose and on the skin, respectively. Therefore, emphasis on controlling overgrowth of the bowel flora by avoidance of agents that raise gastric pH may be relevant only in certain populations, such as liver transplant recipients and patients who have undergone other major intraabdominal procedures or who have bowel obstruction.

In outbreaks of VAP due to specific pathogens, the possibility of a breakdown in infection control measures (particularly contamination of reusable equipment) should be investigated. Even high rates of pathogens that are already common in a particular ICU may result from cross-infection. Education and reminders of the need for consistent hand washing and other infection-control practices can minimize this risk.

HOSPITAL-ACQUIRED PNEUMONIA

While significantly less well studied than VAP, HAP in non-intubated patients—both inside and outside the ICU—is similar to VAP. The main differences are the higher frequency of non-MDR pathogens and the generally better underlying host immunity in non-intubated patients. The lower frequency of MDR pathogens allows monotherapy in a larger proportion of cases of HAP than of VAP.

The only pathogens that may be more common in the non-VAP population are anaerobes. The greater risk of macroaspiration by non-intubated patients and the lower oxygen tensions in the lower respiratory tract of these patients increase the likelihood of a role for anaerobes. While more common in patients with HAP, anaerobes usually contribute only to polymicrobial pneumonias. As in the management of CAP, specific therapy targeting anaerobes probably is not needed since many of the recommended antibiotics are active against anaerobes.

Diagnosis is even more difficult for HAP in the non-intubated patient than for VAP. Lower respiratory tract samples appropriate for culture are considerably more difficult to obtain from non-intubated patients. Many of the underlying diseases that predispose a patient to HAP are also associated with an inability to cough adequately. Since blood cultures are infrequently positive (<15% of cases), the majority of patients with HAP do not have culture data on which antibiotic modifications can be based. Therefore, de-escalation of therapy is less likely in patients with risk factors for MDR pathogens. Despite these difficulties, the better host defenses in non-ICU patients result in lower mortality rates than are documented for VAP. In addition, the risk of antibiotic failure is lower in HAP.

GLOBAL IMPACT



From the available data, it is virtually impossible to accurately assess the impact of pneumonia from a global perspective. Any differences in incidence, disease burden, and costs across different age, ethnic, and racial groups are compounded by differences among countries in terms of etiologic pathogens, resistance rates, access to health-care and diagnostic facilities, and vaccine availability and usage.

A standard approach with clearly defined outcome measures is needed before the impact of pneumonia can be accurately evaluated. However, simple extrapolation from U.S. data for CAP and HAP/VAP shows that pneumonia has a significant impact on quality of life, morbidity, health costs, and mortality rates and that this impact has implications for patients and for society as a whole.

FURTHER READING

Chastre J et al: Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial. JAMA 290:2588, 2003. [[PubMed: 14625336](#)]

Fagon JY et al: Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia. A randomized trial. Ann Intern Med 132:621, 2000. [[PubMed: 10766680](#)]

Jain S et al: Community-acquired pneumonia requiring hospitalization among U.S. adults. N Engl J Med 373:415, 2015. [[PubMed: 26172429](#)]

Kalil AC et al: Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis 63:e61, 2016. [[PubMed: 27418577](#)]

Mandell LA et al: Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis 44(Suppl 2):S27, 2007. [[PubMed: 17278083](#)]

Musher DM, Thorner AR: Community-acquired pneumonia. N Engl J Med 371:1619, 2014. [[PubMed: 25337751](#)]

Shindo Y et al: Risk factors for drug-resistant pathogens in community-acquired and healthcare-associated pneumonia. Am J Respir Crit Care Med 188:985, 2013. [[PubMed: 23855620](#)]

Wunderink RG, Waterer GW: Community-acquired pneumonia. N Engl J Med 370:541, 2014.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 122: Lung Abscess

Rebecca M. Baron; Miriam Baron Barshak

FIGURE 122-1

INTRODUCTION

Lung abscess represents necrosis and cavitation of the lung following microbial infection. Lung abscesses can be single or multiple but usually are marked by a single dominant cavity >2 cm in diameter.

ETIOLOGY

The low prevalence of lung abscesses makes them difficult to study in randomized controlled trials. Although the incidence of lung abscesses has decreased in the antibiotic era, they are still a source of significant morbidity and mortality.

Lung abscesses are usually characterized as either primary (~80% of cases) or secondary. *Primary* lung abscesses usually arise from aspiration, are often caused principally by anaerobic bacteria, and occur in the absence of an underlying pulmonary or systemic condition. *Secondary* lung abscesses arise in the setting of an underlying condition, such as a postobstructive process (e.g., a bronchial foreign body or tumor) or a systemic process (e.g., HIV infection or another immunocompromising condition). Lung abscesses can also be characterized as acute (<4–6 weeks in duration) or chronic (~40% of cases).

EPIDEMIOLOGY

The majority of the existing epidemiologic information involves primary lung abscesses. In general, middle-aged men are more commonly affected than middle-aged women. The major risk factor for primary lung abscesses is aspiration. Patients at particular risk for aspiration, such as those with altered mental status, alcoholism, drug overdose, seizures, bulbar dysfunction, prior cerebrovascular or cardiovascular events, or neuromuscular disease, are most commonly affected. In addition, patients with esophageal dysmotility or esophageal lesions (strictures or tumors) and those with gastric distention and/or gastroesophageal reflux, especially those who spend substantial time in the recumbent position, are at risk for aspiration.

It is widely thought that colonization of the gingival crevices by anaerobic bacteria or microaerophilic streptococci (especially in patients with gingivitis and periodontal disease), combined with a risk of

aspiration, is important in the development of lung abscesses. In fact, many physicians consider it extremely rare for lung abscesses to develop in the absence of teeth as a nidus for bacterial colonization.

The importance of these risk factors in the development of lung abscesses is highlighted by a significant reduction in abscess incidence in the late 1940s that coincided with a change in oral surgical technique: beginning at that time, these operations were no longer performed with the patient in the seated position without a cuffed endotracheal tube, and the frequency of perioperative aspiration events was thus decreased. In addition, the introduction of penicillin around the same time significantly reduced the incidence of and mortality rate from lung abscess.

PATHOGENESIS

Primary Lung Abscesses

The development of primary lung abscesses is thought to originate when chiefly anaerobic bacteria (as well as microaerophilic streptococci) in the gingival crevices are aspirated into the lung parenchyma in a susceptible host (**Table 122-1**). Patients who develop primary lung abscesses usually carry an overwhelming burden of aspirated material or are unable to clear the bacterial load. Pneumonitis develops initially (exacerbated in part by tissue damage caused by gastric acid); then, over a period of 7–14 days, the anaerobic bacteria produce parenchymal necrosis and cavitation whose extent depends on host–pathogen interaction (**Fig. 122-1**). Anaerobes are thought to produce more extensive tissue necrosis in polymicrobial infections in which virulence factors of the various bacteria can act synergistically to cause more significant tissue destruction.

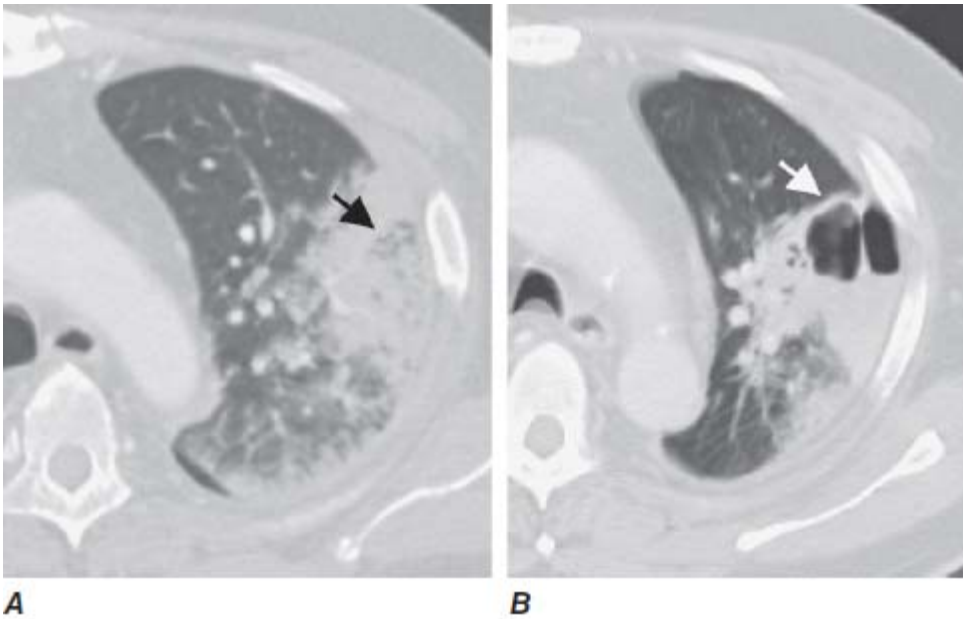
TABLE 122-1

Examples of Microbial Pathogens That Can Cause Lung Abscesses

CLINICAL CONDITION	PATHOGENS
Primary lung abscess (usually with risk factors for aspiration)	Anaerobes (e.g., <i>Peptostreptococcus</i> spp., <i>Prevotella</i> spp., <i>Bacteroides</i> spp., <i>Streptococcus milleri</i>), microaerophilic streptococci
Secondary lung abscess (often with underlying immunocompromise)	<i>Staphylococcus aureus</i> , gram-negative rods (e.g., <i>Pseudomonas aeruginosa</i> , Enterobacteriaceae), <i>Nocardia</i> spp., <i>Aspergillus</i> spp., Mucorales, <i>Cryptococcus</i> spp., <i>Legionella</i> spp., <i>Rhodococcus equi</i> , <i>Pneumocystis jirovecii</i>
Embolic lesions	<i>Staphylococcus aureus</i> (often from endocarditis), <i>Fusobacterium necrophorum</i> (Lemierre's syndrome; see text for details)
Endemic infections (with or without underlying immunocompromise)	<i>Mycobacterium tuberculosis</i> (as well as <i>Mycobacterium avium</i> and <i>Mycobacterium kansasii</i>), <i>Coccidioides</i> spp., <i>Histoplasma capsulatum</i> , <i>Blastomyces</i> spp., parasites (e.g., <i>Entamoeba histolytica</i> , <i>Paragonimus westermani</i> , <i>Strongyloides stercoralis</i>)
Miscellaneous conditions	Bacterial pathogen (often <i>S. aureus</i>) after influenza or another viral infection, <i>Actinomyces</i> spp.

FIGURE 122-1

Representative chest CT scans demonstrating development of lung abscesses. This patient was immunocompromised by underlying lymphoma and developed severe *Pseudomonas aeruginosa* pneumonia, as represented by a left lung infiltrate with concern for central regions of necrosis (*panel A*, *black arrow*). Two weeks later, areas of cavitation with air-fluid levels were visible in this region and were consistent with the development of lung abscesses (*panel B*, *white arrow*). (Images provided by Dr. Ritu Gill, Division of Chest Radiology, Brigham and Women's Hospital, Boston; with permission.)

**A****B**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Secondary Lung Abscesses

The pathogenesis of secondary abscesses depends on the predisposing factor. For example, in cases of bronchial obstruction from malignancy or a foreign body, the obstructing lesion prevents clearance of oropharyngeal secretions, leading to abscess development. With underlying systemic conditions (e.g., immunosuppression after bone marrow or solid organ transplantation), impaired host defense mechanisms lead to increased susceptibility to the development of lung abscesses caused by a broad range of pathogens, including opportunistic organisms (Table 122-1).

Lung abscesses also arise from septic emboli, either in tricuspid valve endocarditis (often involving *Staphylococcus aureus*) or in Lemierre's syndrome, in which an infection begins in the pharynx (classically involving *Fusobacterium necrophorum*) and then spreads to the neck and the carotid sheath (which contains the jugular vein) to cause septic thrombophlebitis.

PATHOLOGY AND MICROBIOLOGY

Primary Lung Abscesses

The dependent segments (posterior upper lobes and superior lower lobes) are the most common locations of primary lung abscesses, given the predisposition of aspirated materials to be deposited in these areas. Generally, the right lung is affected more commonly than the left because the right mainstem bronchus is less angulated.

The microbiology of primary lung abscesses is often polymicrobial, primarily including anaerobic organisms as well as microaerophilic streptococci (Table 122-1). The retrieval and culture of anaerobes can be complicated by the contamination of samples with microbes from the oral cavity, the need for expeditious

transport of the cultures to the laboratory, the need for early plating with special culture techniques, the prolonged time required for culture growth, and the need for collection of specimens prior to administration of antibiotics. When attention is paid to these factors, rates of recovery of specific isolates are reportedly as high as 78%.

Because it is not clear that knowing the identity of the causative anaerobic isolate alters the response to treatment of a primary lung abscess, practice has shifted away from the use of specialized techniques to obtain material for culture, such as transtracheal aspiration and bronchoalveolar lavage with protected brush specimens that allow recovery of culture material while avoiding contamination from the oral cavity. When no pathogen is isolated from a primary lung abscess (which is the case as often as 40% of the time), the abscess is termed a *nonspecific lung abscess*, and the presence of anaerobes is often presumed. A *putrid lung abscess* refers to cases with foul-smelling breath, sputum, or empyema; these manifestations are essentially diagnostic of an anaerobic lung abscess.

Secondary Lung Abscesses

The location of secondary abscesses may vary with the underlying cause. The microbiology of secondary lung abscesses can encompass quite a broad bacterial spectrum, with infection by *Pseudomonas aeruginosa* and other gram-negative rods most common. In addition, a broad array of pathogens can be identified in patients from certain endemic areas and in specific clinical scenarios (e.g., a significant incidence of fungal infections among immunosuppressed patients following bone marrow or solid organ transplantation). Because immunocompromised hosts and patients without the classic presentation of a primary lung abscess can be infected with a wide array of unusual organisms ([Table 122-1](#)), it is of special importance to obtain culture material in order to target therapy.

CLINICAL MANIFESTATIONS

Clinical manifestations may initially be similar to those of pneumonia, with fevers, cough, sputum production, and chest pain; a more chronic and indolent presentation that includes night sweats, fatigue, and anemia is often observed with anaerobic lung abscesses. A subset of patients with putrid lung abscesses may report discolored phlegm and foul-tasting or foul-smelling sputum. Patients with lung abscesses due to non-anaerobic organisms, such as *S. aureus*, may present with a more fulminant course characterized by high fevers and rapid progression.

Findings on physical examination may include fevers, poor dentition, and/or gingival disease as well as amphoric and/or cavernous breath sounds on lung auscultation. Additional findings may include digital clubbing and the absence of a gag reflex.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of lung abscesses is broad and includes other noninfectious processes that result in cavitary lung lesions, including lung infarction, malignancy, sequestration, cryptogenic organizing pneumonia, sarcoidosis, vasculitides and other autoimmune diseases (e.g., granulomatosis with polyangiitis), lung cysts or bullae containing fluid, and septic emboli (e.g., from tricuspid valve endocarditis). Other less common entities can include pulmonary manifestations of diseases that usually present at locations other than the chest (e.g., inflammatory bowel disease, pyoderma gangrenosum).

DIAGNOSIS

Lung abscesses are documented by chest imaging. Although a chest radiograph usually detects a thick-walled cavity with an air-fluid level, CT permits better definition and may provide earlier evidence of cavitation. CT may also yield additional information regarding a possible underlying cause of lung abscess, such as malignancy, and may help distinguish a peripheral lung abscess from a pleural infection. This distinction has important implications for treatment because a pleural space infection, such as an empyema, may require urgent drainage.

As described earlier (see “[Pathology and Microbiology](#),” above), more invasive diagnostics (such as transtracheal aspiration) were traditionally undertaken for primary lung abscesses, whereas empirical therapy that includes drugs targeting anaerobic organisms currently is used more often. While sputum can be collected noninvasively for Gram’s stain and culture, which may yield a pathogen, the infection is likely to be polymicrobial, and culture results may not reflect the presence of anaerobic organisms. As stated above, many physicians consider putrid-smelling sputum to be virtually diagnostic of an anaerobic infection.

When a secondary lung abscess is present or empirical therapy fails to elicit a response, sputum and blood cultures are advised in addition to serologic studies for opportunistic pathogens (e.g., viruses and fungi causing infections in immunocompromised hosts). Additional diagnostics, such as bronchoscopy with bronchoalveolar lavage or protected brush specimen collection and CT-guided percutaneous needle aspiration, can be undertaken. Risks posed by these more invasive diagnostics include spillage of abscess contents into the other lung (with bronchoscopy) and pneumothorax and bronchopleural fistula development (with CT-guided needle aspiration). However, early diagnostics in secondary abscesses, especially in immunocompromised hosts, are particularly important, because the patients involved may be especially fragile, at risk for infection with a broad array of pathogens, and therefore less likely than other patients to respond to empirical therapy.

TREATMENT

TREATMENT

Lung Abscess

The availability of antibiotics in the 1940s and 1950s established therapy with this drug class as the primary approach to the treatment of lung abscess. Previously, surgery had been relied upon much more frequently.

For many decades, penicillin was the antibiotic of choice for primary lung abscesses in light of its anaerobic coverage; however, because oral anaerobes can produce β -lactamases, [clindamycin](#) has proved superior to penicillin in clinical trials. For primary lung abscesses, the recommended regimens are (1) [clindamycin](#) (600 mg IV three times daily; then, with the disappearance of fever and clinical improvement, 300 mg PO four times daily) or (2) an IV-administered β -lactam/ β -lactamase combination, followed—once the patient's condition is stable—by orally administered amoxicillin-clavulanate. This therapy should be continued until imaging demonstrates that the lung abscess has cleared or regressed to a small scar. Treatment duration may range from 3–4 weeks to as long as 14 weeks. One small study suggested that [moxifloxacin](#) (400 mg/d PO) is as effective and well tolerated as ampicillin-sulbactam. Notably, [metronidazole](#) is not effective as a single agent: it covers anaerobic organisms but not the microaerophilic streptococci that are often components of the mixed flora of primary lung abscesses.

In secondary lung abscesses, antibiotic coverage should be directed at the identified pathogen, and a prolonged course (until resolution of the abscess is documented) is often required. Treatment regimens and courses vary widely, depending on the immune state of the host and the identified pathogen. Other interventions may be necessary as well, such as relief of an obstructing lesion or treatment directed at the underlying condition predisposing the patient to lung abscess. Similarly, if the condition of patients with presumed primary lung abscess fails to improve, additional studies to rule out an underlying predisposing cause for a secondary lung abscess are indicated.

Although it can take as long as 7 days for patients receiving appropriate therapy to defervesce, as many as 10–20% of patients may not respond at all, with continued fevers and progression of the abscess cavity on imaging. An abscess >6–8 cm in diameter is less likely to respond to antibiotic therapy without additional interventions. Options for patients who do not respond to antibiotics and whose additional diagnostic studies fail to identify an additional pathogen that can be treated include surgical resection and percutaneous drainage of the abscess, especially when the patient is a poor surgical candidate. Timing of surgical intervention can be challenging; the goal is to balance the morbidity/mortality risk of a procedure with the need for definitively clearing the abscess in the setting of persistent infection that is not responsive to nonsurgical approaches. Possible complications of percutaneous drainage include bacterial contamination of the pleural space as well as pneumothorax and hemothorax.

COMPLICATIONS

Larger cavity size on presentation may correlate with the development of persistent cystic changes (pneumatocoles) or bronchiectasis. Additional possible complications include recurrence of abscesses despite appropriate therapy, extension to the pleural space with development of empyema, life-threatening hemoptysis, and massive aspiration of lung abscess contents.

PROGNOSIS AND PREVENTION

Reported mortality rates for primary abscesses have been as low as 2%, while rates for secondary abscesses are generally higher—as high as 75% in some case series. Other poor prognostic factors include an age of >60, the presence of aerobic bacteria, sepsis at presentation, symptom duration of >8 weeks, and abscess size of >6 cm.

Mitigation of underlying risk factors may be the best approach to prevention of lung abscesses, with attention directed toward airway protection, oral hygiene, and minimized sedation with elevation of the head of the bed for patients at risk for aspiration. Prophylaxis against certain pathogens in at-risk patients (e.g., recipients of bone marrow or solid organ transplants or patients whose immune systems are significantly compromised by HIV infection) may be undertaken.

APPROACH TO THE PATIENT

APPROACH TO THE PATIENT

Lung Abscess

For patients with a lung abscess and a low likelihood of malignancy (e.g., smokers <45 years old) and with risk factors for aspiration, it is reasonable to administer empirical treatment and then to pursue further evaluation if therapy does not elicit a response. However, some clinicians may opt for up-front cultures, even in primary lung abscesses. In patients with risk factors for malignancy or other underlying conditions (especially immunocompromised hosts) or with an atypical presentation, earlier diagnostics should be considered, such as bronchoscopy with biopsy or CT-guided needle aspiration. Bronchoscopy should be performed early in patients whose history, symptoms, or imaging findings are consistent with possible bronchial obstruction. In patients from areas endemic for tuberculosis or patients with other risk factors for tuberculosis (e.g., underlying HIV infection), induced sputum samples should be examined early in the workup to rule out this disease.

FURTHER READING

Bartlett JG: How important are anaerobic bacteria in aspiration pneumonia: When should they be treated and what is optimal therapy. *Infect Dis Clin North Am* 27:149, 2013. [[PubMed: 23398871](#)]

Davis B, Systrom DM: Lung abscess: Pathogenesis, diagnosis, and treatment. *Curr Clin Top Infect Dis* 18:252, 1998. [[PubMed: 9779359](#)]

Desai H, Agrawal A: Pulmonary emergencies: Pneumonia, acute respiratory distress syndrome, lung abscess, and empyema. *Med Clin North Am* 96:1127, 2012. [[PubMed: 23102481](#)]

Ott SR et al: [Moxifloxacin](#) vs ampicillin/sulbactam in aspiration pneumonia and primary lung abscess. *Infection* 36:23, 2008. [[PubMed: 18231720](#)]

Raymond D: Surgical intervention for thoracic infections. Surg Clin North Am 94:1283, 2014. [[PubMed: 25440124](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 124: Infections of the Skin, Muscles, and Soft Tissues

Dennis L. Stevens

INTRODUCTION



Skin and soft tissue infections occur in all races, all ethnic groups, and all geographic locations, although some have unique geographic niches. In modern times, the frequency and severity of some skin and soft tissue infections have increased for several reasons. First, microbes are rapidly disseminated throughout the world via efficient air travel, acquiring genes for virulence factors and antibiotic resistance. Second, natural disasters, such as earthquakes, tsunamis, tornadoes, and hurricanes, appear to be increasing in frequency, and the injuries sustained during these events commonly cause major skin and soft-tissue damage that predisposes to infection. Third, trauma and casualties resulting from combat and terrorist activities can markedly damage or destroy tissues and provide both endogenous and exogenous pathogens with ready access to deeper structures. Unfortunately, because the marvels of modern medicine may not be available during human-instigated and natural disasters, primary treatment may be delayed and the likelihood of severe infection and death increased.

ANATOMIC RELATIONSHIPS: CLUES TO THE DIAGNOSIS OF SOFT TISSUE INFECTIONS

Skin and soft tissue infections have been common human afflictions for centuries. However, between 2000 and 2004, hospital admissions for these infections rose by 27%, a remarkable increase that was attributable largely to the emergence of the USA300 clone of methicillin-resistant *Staphylococcus aureus* (MRSA). This chapter provides an anatomic approach to understanding the types of soft tissue infections and the diverse microbes responsible.

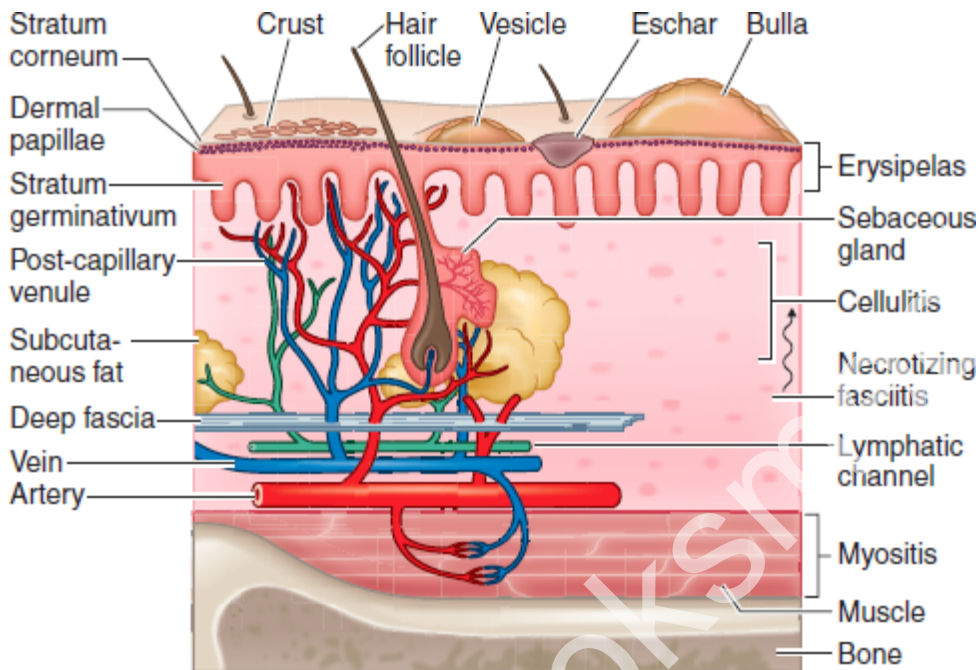
Protection against infection of the epidermis depends on the mechanical barrier afforded by the stratum corneum, since the epidermis itself is devoid of blood vessels (**Fig. 124-1**). Disruption of this layer by burns or bites, abrasions, foreign bodies, primary dermatologic disorders (e.g., herpes simplex, varicella, ecthyma gangrenosum), surgery, or vascular or pressure ulcer allows penetration of bacteria to the deeper structures. Similarly, the hair follicle can serve as a portal either for components of the normal flora (e.g., *Staphylococcus*) or for extrinsic bacteria (e.g., *Pseudomonas* in hot-tub folliculitis). Intracellular infection of the squamous epithelium with vesicle formation may arise from cutaneous inoculation, as in infection with

herpes simplex virus (HSV) type 1; from the dermal capillary plexus, as in varicella and infections due to other viruses associated with viremia; or from cutaneous nerve roots, as in herpes zoster. Bacteria infecting the epidermis, such as *Streptococcus pyogenes*, may be translocated laterally to deeper structures via lymphatics, an event that results in the rapid superficial spread of erysipelas. Later, engorgement or obstruction of lymphatics causes flaccid edema of the epidermis, another characteristic of erysipelas.

FIGURE 124-1

Structural components of the skin and soft tissues, superficial infections, and infections of the deeper structures.

The rich capillary network beneath the dermal papillae plays a key role in the localization of infection and in the development of the acute inflammatory reaction.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The rich plexus of capillaries beneath the dermal papillae provides nutrition to the stratum germinativum, and physiologic responses of this plexus produce important clinical signs and symptoms. For example, infective vasculitis of the plexus results in petechiae, Osler's nodes, Janeway lesions, and palpable purpura, which, if present, are important clues to the existence of endocarditis (**Chap. 123**). In addition, metastatic infection within this plexus can result in cutaneous manifestations of disseminated fungal infection (**Chap. 211**), gonococcal infection (**Chap. 151**), *Salmonella* infection (**Chap. 160**), *Pseudomonas* infection (i.e., ecthyma gangrenosum; **Chap. 159**), meningococcemia (**Chap. 150**), and staphylococcal infection (**Chap. 142**). The plexus also provides bacteria with access to the circulation, thereby facilitating local spread or bacteremia. The postcapillary venules of this plexus are a prominent site of polymorphonuclear leukocyte sequestration, diapedesis, and chemotaxis to the site of cutaneous infection.

Amplification of these physiologic mechanisms by excessive levels of cytokines or bacterial toxins causes leukostasis, venous occlusion, and pitting edema. Edema with purple bullae, ecchymosis, and cutaneous anesthesia suggests loss of vascular integrity and necessitates exploration of the deeper structures for evidence of necrotizing fasciitis or myonecrosis. An early diagnosis requires a high level of suspicion in

instances of unexplained fever and of pain and tenderness in the soft tissue, even in the absence of acute cutaneous inflammation.

Table 124-1 indicates the chapters in which the infections described below are discussed in greater detail. Many of these infections are illustrated in the chapters cited or in **Chap. A1**.

TABLE 124-1

Skin and Soft Tissue Infections

LESION, CLINICAL SYNDROME	INFECTIOUS AGENT(S)	SEE ALSO CHAP(S).
Vesicles		
Smallpox	Variola virus	S2
Chickenpox	Varicella-zoster virus	188
Shingles (herpes zoster)	Varicella-zoster virus	188
Cold sores, herpetic whitlow, herpes gladiatorum	Herpes simplex virus	187
Hand-foot-and-mouth disease	Coxsackievirus A16	199
Orf	Parapoxvirus	191
Molluscum contagiosum	Molluscum contagiosum poxvirus	191
Rickettsialpox	<i>Rickettsia akari</i>	182
Blistering distal dactylitis	<i>Staphylococcus aureus</i> or <i>Streptococcus pyogenes</i>	142, 143
Bullae		
Staphylococcal scalded-skin syndrome	<i>S. aureus</i>	142
Necrotizing fasciitis	<i>S. pyogenes</i> , <i>Clostridium</i> spp., mixed aerobes and anaerobes	143, 149, 172
Gas gangrene	<i>Clostridium</i> spp.	149
Halophilic vibrio	<i>Vibrio vulnificus</i>	163
Crusted lesions		

LESION, CLINICAL SYNDROME	INFECTIOUS AGENT(S)	SEE ALSO CHAP(S).
Bullous impetigo/ecthyma	<i>S. aureus</i>	142
Impetigo contagiosa	<i>S. pyogenes</i>	143
Ringworm	Superficial dermatophyte fungi	214
Sporotrichosis	<i>Sporothrix schenckii</i>	214
Histoplasmosis	<i>Histoplasma capsulatum</i>	207
Coccidioidomycosis	<i>Coccidioides immitis</i>	208
Blastomycosis	<i>Blastomyces dermatitidis</i>	209
Cutaneous leishmaniasis	<i>Leishmania</i> spp.	221
Cutaneous tuberculosis	<i>Mycobacterium tuberculosis</i>	173
Nocardiosis	<i>Nocardia asteroides</i>	169
Folliculitis		
Furunculosis	<i>S. aureus</i>	142
Hot-tub folliculitis	<i>Pseudomonas aeruginosa</i>	159
Swimmer's itch	<i>Schistosoma</i> spp.	229
Acne vulgaris	<i>Propionibacterium acnes</i>	53
Papular and nodular lesions		
Fish-tank or swimming-pool granuloma	<i>Mycobacterium marinum</i>	175
Creeping eruption (cutaneous larva migrans)	<i>Ancylostoma braziliense</i>	226
Dracunculiasis	<i>Dracunculus medinensis</i>	228
Cercarial dermatitis	<i>Schistosoma mansoni</i>	229

LESION, CLINICAL SYNDROME	INFECTIOUS AGENT(S)	SEE ALSO CHAP(S).
Verruca vulgaris	Human papillomaviruses 1, 2, 4	193
Condylomata acuminata (anogenital warts)	Human papillomaviruses 6, 11, 16, 18	193
Onchocerciasis nodule	<i>Onchocerca volvulus</i>	228
Cutaneous myiasis	<i>Dermatobia hominis</i>	452
Verruca peruana	<i>Bartonella bacilliformis</i>	167
Cat-scratch disease	<i>Bartonella henselae</i>	167
Lepromatous leprosy	<i>Mycobacterium leprae</i>	174
Secondary syphilis (papulosquamous and nodular lesions, condylomata lata)	<i>Treponema pallidum</i>	177
Tertiary syphilis (nodular gummatous lesions)	<i>T. pallidum</i>	177
Ulcers with or without eschars		
Anthrax	<i>Bacillus anthracis</i>	S2
Ulceroglandular tularemia	<i>Francisella tularensis</i>	165, S2
Bubonic plague	<i>Yersinia pestis</i>	166, S2
Buruli ulcer	<i>Mycobacterium ulcerans</i>	175
Leprosy	<i>M. leprae</i>	174
Cutaneous tuberculosis	<i>M. tuberculosis</i>	173
Chancroid	<i>Haemophilus ducreyi</i>	152
Primary syphilis	<i>T. pallidum</i>	177
Erysipelas	<i>S. pyogenes</i>	143

LESION, CLINICAL SYNDROME	INFECTIOUS AGENT(S)	SEE ALSO CHAP(S).
Cellulitis	<i>Staphylococcus</i> spp., <i>Streptococcus</i> spp., various other bacteria	Various
Necrotizing fasciitis		
Streptococcal gangrene	<i>S. pyogenes</i>	143
Fournier's gangrene	Mixed aerobic and anaerobic bacteria	172
Staphylococcal necrotizing fasciitis	Methicillin-resistant <i>S. aureus</i>	142
Myositis and myonecrosis		
Pyomyositis	<i>S. aureus</i>	142
Streptococcal necrotizing myositis	<i>S. pyogenes</i>	143
Gas gangrene	<i>Clostridium</i> spp.	149
Nonclostridial (crepitant) myositis	Mixed aerobic and anaerobic bacteria	172
Synergistic nonclostridial anaerobic myonecrosis	Mixed aerobic and anaerobic bacteria	172

INFECTIONS ASSOCIATED WITH VESICLES

(Table 124-1) Vesicle formation due to infection is caused by viral proliferation within the epidermis. In varicella and variola, viremia precedes the onset of a diffuse centripetal rash that progresses from macules to vesicles, then to pustules, and finally to scabs over the course of 1–2 weeks. Vesicles of varicella have a “dewdrop” appearance and develop in crops randomly about the trunk, extremities, and face over 3–4 days. Herpes zoster occurs in a single dermatome; the appearance of vesicles is preceded by pain for several days. Zoster may occur in persons of any age but is most common among immunosuppressed individuals and elderly patients, whereas most cases of varicella occur in young children. Vesicles due to HSV are found on or around the lips (HSV-1) or genitals (HSV-2) but also may appear on the head and neck of young wrestlers (herpes gladiatorum) or on the digits of health care workers (herpetic whitlow). Recurrent herpes labialis (HSV-1) and herpes genitalis commonly follow primary infection. Coxsackievirus A16 characteristically causes vesicles on the hands, feet, and mouth of children. Orf is caused by a DNA virus related to smallpox virus and

infects the fingers of individuals who work around goats and sheep. Molluscum contagiosum virus induces flaccid vesicles on the skin of healthy and immunocompromised individuals. Although variola (smallpox) in nature was eradicated as of 1977, postmillennial terrorist events have renewed interest in this devastating infection (**Chap. S2**). Viremia beginning after an incubation period of 12 days is followed by a diffuse maculopapular rash, with rapid evolution to vesicles, pustules, and then scabs. Secondary cases can occur among close contacts.

Rickettsialpox begins after mite-bite inoculation of *Rickettsia akari* into the skin. A papule with a central vesicle evolves to form a 1- to 2.5-cm painless crusted black eschar with an erythematous halo and proximal adenopathy. While more common in the northeastern United States and Ukraine in 1940–1950, rickettsialpox has recently been described in Ohio, Arizona, and Utah. Blistering dactylitis is a painful, vesicular, localized *S. aureus* or group A streptococcal infection of the pulps of the distal digits of the hands.

INFECTIONS ASSOCIATED WITH BULLAE

(**Table 124-1**) Staphylococcal scalded-skin syndrome (SSSS) in neonates is caused by a toxin (exfoliatin) from phage group II *S. aureus*. SSSS must be distinguished from toxic epidermal necrolysis (TEN), which occurs primarily in adults, is drug-induced, and is associated with a higher mortality rate. Punch biopsy with frozen section is useful in making this distinction since the cleavage plane is the stratum corneum in SSSS and the stratum germinativum in TEN (**Fig. 124-1**). Intravenous γ -globulin is a promising treatment for TEN. Necrotizing fasciitis and gas gangrene also induce bulla formation (see “**Necrotizing Fasciitis**,” below). Halophilic vibrio infection can be as aggressive and fulminant as necrotizing fasciitis; a helpful clue in its diagnosis is a history of exposure to waters of the Gulf of Mexico or the Atlantic seaboard or (in a patient with cirrhosis) the ingestion of raw seafood. The etiologic organism (*Vibrio vulnificus*) is highly susceptible to tetracycline.

INFECTIONS ASSOCIATED WITH CRUSTED LESIONS

(**Table 124-1**) Impetigo contagiosa is caused by *S. pyogenes*, and bullous impetigo is due to *S. aureus*. Both skin lesions may have an early bullous stage but then appear as thick crusts with a golden-brown color. Epidemics of impetigo caused by MRSA have been reported. Streptococcal lesions are most common among children 2–5 years of age, and epidemics may occur in settings of poor hygiene, particularly among children in lower socioeconomic settings in tropical climates. It is important to recognize impetigo contagiosa because of its relationship to poststreptococcal glomerulonephritis. Rheumatic fever is not a complication of skin infection caused by *S. pyogenes*. Superficial dermatophyte infection (ringworm) can occur on any skin surface, and skin scrapings with KOH staining are diagnostic. Primary infections with dimorphic fungi such as *Blastomyces dermatitidis* and *Sporothrix schenckii* can initially present as crusted skin lesions resembling ringworm. Disseminated infection with *Coccidioides immitis* can also involve the skin, and biopsy and culture should be performed on crusted lesions when the patient is from an endemic area. Crusted nodular lesions caused by *Mycobacterium chelonae* have been described in HIV-seropositive patients. Treatment with clarithromycin looks promising.

FOLLICULITIS

(Table 124-1) Hair follicles serve as portals for a number of bacteria, although *S. aureus* is the most common cause of localized folliculitis. Sebaceous glands empty into hair follicles and ducts and, if these portals are blocked, form sebaceous cysts that may resemble staphylococcal abscesses or may become secondarily infected. Inflammation of sweat glands (hidradenitis suppurativa) also can mimic infection of hair follicles, particularly in the axillae, but new treatments with potent anti-inflammatory agents hold promise. Chronic folliculitis is uncommon except in acne vulgaris, where constituents of the normal flora (e.g., *Propionibacterium acnes*) may play a role.

Diffuse folliculitis occurs in two settings. *Hot-tub folliculitis* is caused by *Pseudomonas aeruginosa* in waters that are insufficiently chlorinated and maintained at temperatures of 37–40°C. Infection is usually self-limited, although bacteremia and shock have been reported. *Swimmer's itch* occurs when a skin surface is exposed to water infested with freshwater avian schistosomes. Warm water temperatures and alkaline pH are suitable for mollusks that serve as intermediate hosts between birds and humans. Free-swimming schistosomal cercariae readily penetrate human hair follicles or pores but quickly die and elicit a brisk allergic reaction, causing intense itching and erythema.

PAPULAR AND NODULAR LESIONS

(Table 124-1) Raised lesions of the skin occur in many different forms. *Mycobacterium marinum* infections of the skin may present as cellulitis or as raised erythematous nodules. Similar lesions caused by *Mycobacterium abscessus* and *M. chelonae* have been described among patients undergoing cosmetic laser surgery and tattooing, respectively. Erythematous papules are early manifestations of cat-scratch disease (with lesions developing at the primary site of inoculation of *Bartonella henselae*) and bacillary angiomatosis (also caused by *B. henselae*). Raised serpiginous or linear eruptions are characteristic of cutaneous larva migrans, which is caused by burrowing larvae of dog or cat hookworms (*Ancylostoma braziliense*) and which humans acquire through contact with soil that has been contaminated with dog or cat feces. Similar burrowing raised lesions are present in dracunculiasis caused by migration of the adult female nematode *Dracunculus medinensis*. Nodules caused by *Onchocerca volvulus* measure 1–10 cm in diameter and occur mostly in persons bitten by *Simulium* flies in Africa. The nodules contain the adult worm encased in fibrous tissue. Migration of microfilariae into the eyes may result in blindness. Verruga peruana is caused by *Bartonella bacilliformis*, which is transmitted to humans by the sandfly *Phlebotomus*. This condition can take the form of single gigantic lesions (several centimeters in diameter) or multiple small lesions (several millimeters in diameter). Numerous subcutaneous nodules may also be present in cysticercosis caused by larvae of *Taenia solium*. Multiple erythematous papules develop in schistosomiasis; each represents a cercarial invasion site. Skin nodules as well as thickened subcutaneous tissue are prominent features of lepromatous leprosy. Large nodules or gummas are features of tertiary syphilis, whereas flat papulosquamous lesions are characteristic of secondary syphilis. Human papillomavirus may cause singular warts (verruca vulgaris) or multiple warts in the anogenital area (condylomata acuminata). The latter are major problems in HIV-infected individuals.

ULCERS WITH OR WITHOUT ESCHARS

(Table 124-1) Cutaneous anthrax begins as a pruritic papule, which develops within days into an ulcer with surrounding vesicles and edema and then into an enlarging ulcer with a black eschar. Cutaneous anthrax may cause chronic nonhealing ulcers with an overlying dirty-gray membrane, although lesions may also mimic psoriasis, eczema, or impetigo. Ulceroglandular tularemia may have associated ulcerated skin lesions with painful regional adenopathy. Although buboes are the major cutaneous manifestation of plague, ulcers with eschars, papules, or pustules are also present in 25% of cases.

Mycobacterium ulcerans typically causes chronic skin ulcers on the extremities of individuals living in the tropics. *Mycobacterium leprae* may be associated with cutaneous ulcerations in patients with lepromatous leprosy related to Lucio's phenomenon, in which immune-mediated destruction of tissue bearing high concentrations of *M. leprae* bacilli occurs, usually several months after initiation of effective therapy. *Mycobacterium tuberculosis* also may cause ulcerations, papules, or erythematous macular lesions of the skin in both immunocompetent and immunocompromised patients.

Decubitus ulcers are due to tissue hypoxemia secondary to pressure-induced vascular insufficiency and may become secondarily infected with components of the skin and gastrointestinal flora, including anaerobes. Ulcerative lesions on the anterior shins may be due to pyoderma gangrenosum, which must be distinguished from similar lesions of infectious etiology by histologic evaluation of biopsy sites. Ulcerated lesions on the genitals may be either painful (chancroid) or painless (primary syphilis).

ERYSIPELAS

(Table 124-1) Erysipelas is due to *S. pyogenes* and is characterized by an abrupt onset of fiery-red swelling of the face or extremities. The distinctive features of erysipelas are well-defined indurated margins, particularly along the nasolabial fold; rapid progression; and intense pain. Flaccid bullae may develop during the second or third day of illness, but extension to deeper soft tissues is rare. Treatment with penicillin is effective; swelling may progress despite appropriate treatment, although fever, pain, and the intense red color diminish. Desquamation of the involved skin occurs 5–10 days into the illness. Infants and elderly adults are most commonly afflicted, and the severity of systemic toxicity varies.

CELLULITIS

(Table 124-1) Cellulitis is an acute inflammatory condition of the skin that is characterized by localized pain, erythema, swelling, and heat. It may be caused by indigenous flora colonizing the skin and appendages (e.g., *S. aureus* and *S. pyogenes*) or by a wide variety of exogenous bacteria. Because the exogenous bacteria involved in cellulitis occupy unique niches in nature, a thorough history (including epidemiologic data) offers important clues to etiology. When there is drainage, an open wound, or an obvious portal of entry, Gram's stain and culture provide a definitive diagnosis. In the absence of these findings, the bacterial etiology of cellulitis is difficult to establish, and in some cases staphylococcal and streptococcal cellulitis may have

similar features. Even with needle aspiration of the leading edge or a punch biopsy of the cellulitis tissue itself, cultures are positive in only 20% of cases. This observation suggests that relatively low numbers of bacteria may cause cellulitis and that the expanding area of erythema within the skin may be a direct effect of extracellular toxins or of the soluble mediators of inflammation elicited by the host.

Bacteria may gain access to the epidermis through cracks in the skin, abrasions, cuts, burns, insect bites, surgical incisions, and IV catheters. Cellulitis caused by *S. aureus* spreads from a central localized infection, such as an abscess, folliculitis, or an infected foreign body (e.g., a splinter, a prosthetic device, or an IV catheter). MRSA is rapidly replacing methicillin-sensitive *S. aureus* (MSSA) as a cause of cellulitis in both inpatient and outpatient settings. Cellulitis caused by MSSA or MRSA is usually associated with a focal infection, such as a furuncle, a carbuncle, a surgical wound, or an abscess; the U.S. Food and Drug Administration preferentially refers to these types of infection as *purulent cellulitis*. In contrast, cellulitis due to *S. pyogenes* is a more rapidly spreading, diffuse process that is frequently associated with lymphangitis and fever and should be referred to as *nonpurulent cellulitis*. Recurrent streptococcal cellulitis of the lower extremities may be caused by organisms of group A, C, or G in association with chronic venous stasis or with saphenous venectomy for coronary artery bypass surgery. Streptococci also cause recurrent cellulitis among patients with chronic lymphedema resulting from elephantiasis, lymph node dissection, or Milroy disease. Recurrent staphylococcal cutaneous infections are more common among individuals who have eosinophilia and elevated serum levels of IgE (Job syndrome) and among nasal carriers of staphylococci. Cellulitis caused by *Streptococcus agalactiae* (group B *Streptococcus*) occurs primarily in elderly patients and those with diabetes mellitus or peripheral vascular disease. *Haemophilus influenzae* typically causes periorbital cellulitis in children in association with sinusitis, otitis media, or epiglottitis. It is unclear whether this form of cellulitis will (like meningitis) become less common as a result of the impressive efficacy of the *H. influenzae* type b vaccine.

Many other bacteria also cause cellulitis. It is fortunate that these organisms occur in such characteristic settings that a good history provides useful clues to the diagnosis. Cellulitis associated with cat bites and, to a lesser degree, with dog bites is commonly caused by *Pasteurella multocida*, although in the latter case *Staphylococcus intermedius* and *Capnocytophaga canimorsus* also must be considered. Sites of cellulitis and abscesses associated with dog bites and human bites also contain a variety of anaerobic organisms, including *Fusobacterium*, *Bacteroides*, aerobic and anaerobic streptococci, and *Eikenella corrodens*. *Pasteurella* is notoriously resistant to dicloxacillin and nafcillin but is sensitive to all other β -lactam antimicrobial agents as well as to quinolones, tetracycline, and erythromycin. Amoxicillin-clavulanate, ampicillin-sulbactam, and cefoxitin are good choices for the treatment of animal or human bite infections. *Aeromonas hydrophila* causes aggressive cellulitis and occasionally necrotizing fasciitis in tissues surrounding lacerations sustained in freshwater (lakes, rivers, and streams). This organism remains sensitive to aminoglycosides, fluoroquinolones, chloramphenicol, trimethoprim-sulfamethoxazole, and third-generation cephalosporins; it is resistant to ampicillin, however. *P. aeruginosa* causes three types of soft tissue infection: ecthyma gangrenosum in neutropenic patients, hot-tub folliculitis, and cellulitis following penetrating injury. Most commonly, *P. aeruginosa* is introduced into the deep tissues when a person steps on a nail. Treatment includes surgical inspection and drainage, particularly if the injury also involves bone or

joint capsule. Choices for empirical treatment while antimicrobial susceptibility data are awaited include an aminoglycoside, a third-generation cephalosporin (ceftazidime, cefoperazone, or cefotaxime), a semisynthetic penicillin (ticarcillin, mezlocillin, or piperacillin), or a fluoroquinolone (although drugs of the last class are not indicated for the treatment of children <13 years old).

Gram-negative bacillary cellulitis, including that due to *P. aeruginosa*, is most common among hospitalized, immunocompromised hosts. Cultures and sensitivity tests are critically important in this setting because of multidrug resistance (**Chap. 159**).

The gram-positive aerobic rod *Erysipelothrix rhusiopathiae* is most often associated with fish and domestic swine and causes cellulitis primarily in bone renderers and fishmongers. *E. rhusiopathiae* remains susceptible to most β -lactam antibiotics (including penicillin), erythromycin, clindamycin, tetracycline, and cephalosporins but is resistant to sulfonamides, chloramphenicol, and vancomycin. Its resistance to vancomycin, which is unusual among gram-positive bacteria, is of potential clinical significance since this agent is sometimes used in empirical therapy for skin infection. Fish food containing the water flea *Daphnia* is sometimes contaminated with *M. marinum*, which can cause cellulitis or granulomas on skin surfaces exposed to the water in aquariums or injured in swimming pools. Rifampin plus ethambutol has been an effective therapeutic combination in some cases, although no comprehensive studies have been undertaken. In addition, some strains of *M. marinum* are susceptible to tetracycline or to trimethoprim-sulfamethoxazole.

NECROTIZING FASCIITIS

(**Table 124-1**) Necrotizing fasciitis, formerly called streptococcal gangrene, may be associated with group A *Streptococcus* or mixed aerobic–anaerobic bacteria or may occur as a component of gas gangrene caused by *Clostridium perfringens*. Strains of MRSA that produce the Panton-Valentine leukocidin (PVL) toxin have been reported to cause necrotizing fasciitis. Early diagnosis may be difficult when pain or unexplained fever is the only presenting manifestation. Swelling then develops and is followed by brawny edema and tenderness. With progression, dark-red induration of the epidermis appears, along with bullae filled with blue or purple fluid. Later the skin becomes friable and takes on a bluish, maroon, or black color. By this stage, thrombosis of blood vessels in the dermal papillae (**Fig. 124-1**) is extensive. Extension of infection to the level of the deep fascia causes this tissue to take on a brownish-gray appearance. Rapid spread occurs along fascial planes, through venous channels and lymphatics. Patients in the later stages are toxic and frequently manifest shock and multiorgan failure.

Necrotizing fasciitis caused by mixed aerobic–anaerobic bacteria begins with a breach in the integrity of a mucous membrane barrier, such as the mucosa of the gastrointestinal or genitourinary tract. The portal can be a malignancy, a diverticulum, a hemorrhoid, an anal fissure, or a urethral tear. Other predisposing factors include peripheral vascular disease, diabetes mellitus, surgery, and penetrating injury to the abdomen. Leakage into the perineal area results in a syndrome called *Fournier's gangrene*, characterized by massive swelling of the scrotum and penis with extension into the perineum or the abdominal wall and the legs.

Necrotizing fasciitis caused by *S. pyogenes* has increased in frequency and severity since 1985. There are two distinct clinical presentations: those with no portal of entry and those with a defined portal of entry. Infections in the first category often begin deep at the site of a nonpenetrating minor trauma, such as a bruise or a muscle strain. Seeding of the site via transient bacteremia is likely, although most patients deny antecedent streptococcal infection. The affected patients present with only severe pain and fever. Late in the course, the classic signs of necrotizing fasciitis, such as purple (violaceous) bullae, skin sloughing, and progressive toxicity, develop. In infections of the second type, *S. pyogenes* may reach the deep fascia from a site of cutaneous infection or penetrating trauma. These patients have early signs of superficial skin infection with progression to necrotizing fasciitis. In either case, toxicity is severe, and renal impairment may precede the development of shock. In 20–40% of cases, myositis occurs concomitantly, and, as in gas gangrene (see below), serum creatine phosphokinase levels may be markedly elevated. Necrotizing fasciitis due to mixed aerobic–anaerobic bacteria may be associated with gas in deep tissue, but gas usually is not present when the cause is *S. pyogenes* or MRSA. Prompt surgical exploration down to the deep fascia and muscle is essential. Necrotic tissue must be surgically removed, and Gram's staining and culture of excised tissue are useful in establishing whether group A streptococci, mixed aerobic–anaerobic bacteria, MRSA, or *Clostridium* species are present (see “[Treatment](#),” below).

MYOSITIS AND MYONECROSIS

([Table 124-1](#)) Muscle involvement can occur with viral infection (e.g., influenza, dengue, or coxsackievirus B infection) or parasitic invasion (e.g., trichinellosis, cysticercosis, or toxoplasmosis). Although myalgia develops in most of these infections, severe muscle pain is the hallmark of pleurodynia (coxsackievirus B), trichinellosis, and bacterial infection. Acute rhabdomyolysis predictably occurs with clostridial and streptococcal myositis but may also be associated with influenza virus, echovirus, coxsackievirus, Epstein-Barr virus, and *Legionella* infections.

Pyomyositis is usually due to *S. aureus*, is common in tropical areas, and generally has no known portal of entry. Cases of pyomyositis caused by MRSA producing the PVL toxin have been described among children in the United States. Muscle infection begins at the exact site of blunt trauma or muscle strain. Infection remains localized, and shock does not develop unless organisms produce toxic shock syndrome toxin 1 or certain enterotoxins and the patient lacks antibodies to the toxin produced by the infecting organisms. In contrast, *S. pyogenes* may induce primary myositis (referred to as *streptococcal necrotizing myositis*) in association with severe systemic toxicity. Myonecrosis occurs concomitantly with necrotizing fasciitis in ~50% of cases. Both are part of the streptococcal toxic shock syndrome.

Gas gangrene usually follows severe penetrating injuries that result in interruption of the blood supply and introduction of soil into wounds. Such cases of traumatic gangrene are usually caused by the clostridial species *C. perfringens*, *C. septicum*, and *C. histolyticum*. Rarely, latent or recurrent gangrene can occur years after penetrating trauma; dormant spores that reside at the site of previous injury are most likely responsible. Spontaneous nontraumatic gangrene among patients with neutropenia, gastrointestinal malignancy, diverticulosis, or recent radiation therapy to the abdomen is caused by several clostridial

species, of which *C. septicum* is the most commonly involved. The tolerance of this anaerobe to oxygen probably explains why it can initiate infection spontaneously in normal tissue anywhere in the body.

Gas gangrene of the uterus, especially that due to *Clostridium sordellii*, historically occurred as a consequence of illegal or self-induced abortion and nowadays also follows spontaneous abortion, vaginal delivery, and cesarean section. *C. sordellii* has also been implicated in medically induced abortion.

Postpartum *C. sordellii* infections in young, previously healthy women present as a unique clinical picture: little or no fever, lack of a purulent discharge, refractory hypotension, extensive peripheral edema and effusions, hemoconcentration, and a markedly elevated white blood cell count. The infection is almost uniformly fatal, with death ensuing rapidly. *C. sordellii* and *C. novyi* have also been associated with cutaneous injection of black tar heroin; mortality rates are lower among the affected individuals, probably because their aggressive injection-site infections are readily apparent and diagnosis is therefore prompt.

Synergistic nonclostridial anaerobic myonecrosis, also known as necrotizing cutaneous myositis and synergistic necrotizing cellulitis, is a variant of necrotizing fasciitis caused by mixed aerobic and anaerobic bacteria with the exclusion of clostridial organisms (see “[Necrotizing Fasciitis](#),” above).

DIAGNOSIS

This chapter emphasizes the physical appearance and location of lesions within the soft tissues as important diagnostic clues. Other crucial considerations in narrowing the differential diagnosis are the temporal progression of the lesions as well as the patient’s travel history, animal exposure or bite history, age, underlying disease status, and lifestyle. However, even the astute clinician may find it challenging to diagnose all infections of the soft tissues by history and inspection alone. Soft tissue radiography, CT ([Fig. 124-2](#)), and MRI may be useful in determining the depth of infection and should be performed when the patient has rapidly progressing lesions or evidence of a systemic inflammatory response syndrome. These tests are particularly valuable for defining a localized abscess or detecting gas in tissue. Unfortunately, they may reveal only soft tissue swelling and thus are not specific for fulminant infections such as necrotizing fasciitis or myonecrosis caused by group A *Streptococcus* ([Fig. 124-2](#)), where gas is not found in lesions.

FIGURE 124-2

CT showing edema and inflammation of the left chest wall in a patient with necrotizing fasciitis and myonecrosis caused by group A *Streptococcus*.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Aspiration of the leading edge or punch biopsy with frozen section may be helpful if the results of imaging tests are positive, but false-negative results occur in ~80% of cases. There is some evidence that aspiration alone may be superior to injection and aspiration with normal saline. Frozen sections are especially useful in distinguishing SSSS from TEN and are quite valuable in cases of necrotizing fasciitis. Open surgical inspection, with debridement as indicated, is clearly the best way to determine the extent and severity of infection and to obtain material for Gram's staining and culture. Such an aggressive approach is important and may be lifesaving if undertaken early in the course of fulminant infections where there is evidence of systemic toxicity.

TREATMENT

TREATMENT

Infections of the Skin, Muscles, and Soft Tissues

A full description of the treatment of all the clinical entities described herein is beyond the scope of this chapter. As a guide to the clinician in selecting appropriate treatment, the antimicrobial agents useful in the most common and the most fulminant cutaneous infections are listed in [Table 124-2](#).

Furuncles, carbuncles, and abscesses caused by MRSA and MSSA are common, and their treatment depends upon the size of the lesion. Furuncles <2.5 cm in diameter are usually treated with moist heat. Those that are larger (4.5 cm of erythema and induration) require surgical drainage, and the occurrence of these larger lesions in association with fever, chills, or leukocytosis requires both drainage and antibiotic treatment. Previous studies in children demonstrated that surgical drainage of abscesses (mean diameter, 3.8 cm) was as effective when used alone as when combined with trimethoprim-sulfamethoxazole treatment. However, the rate of recurrence of new lesions was lower in the group undergoing both drainage and antibiotic treatment. Recent studies in adults with predominantly MRSA localized abscesses suggested that a 7- to 10-

day course of treatment with trimethoprim-sulfamethoxazole was associated with higher cure rates and fewer recurrences. In children, a 3-day course was not as effective as a 7-day course.

Early and aggressive surgical exploration is essential in cases of suspected necrotizing fasciitis, myositis, or gangrene in order to (1) visualize the deep structures, (2) remove necrotic tissue, (3) reduce compartment pressure, and (4) obtain suitable material for Gram's staining and for aerobic and anaerobic cultures. Appropriate empirical antibiotic treatment for mixed aerobic-anaerobic infections could consist of ampicillin-sulbactam, cefoxitin, or the following combination: (1) [clindamycin](#) (600–900 mg IV every 8 h) or [metronidazole](#) (500 mg every 6 h) plus (2) ampicillin or ampicillin-sulbactam (1.5–3 g IV every 6 h) plus (3) [gentamicin](#) (1–1.5 mg/kg every 8 h). Group A streptococcal and clostridial infection of the fascia and/or muscle carries a mortality rate of 20–50% with penicillin treatment. In experimental models of streptococcal and clostridial necrotizing fasciitis/myositis, [clindamycin](#) has exhibited markedly superior efficacy, but no comparative clinical trials have been performed. A retrospective study of children with invasive group A streptococcal infection demonstrated higher survival rates with [clindamycin](#) treatment than with β -lactam antibiotic therapy. Hyperbaric oxygen treatment also may be useful in gas gangrene due to clostridial species. Antibiotic treatment should be continued until all signs of systemic toxicity have resolved, all devitalized tissue has been removed, and granulation tissue has developed ([Chaps. 143, 149, and 172](#)).

In summary, infections of the skin and soft tissues are diverse in presentation and severity and offer a great challenge to the clinician. This chapter provides an approach to diagnosis and understanding of the pathophysiologic mechanisms involved in these infections. More in-depth information is found in chapters on specific infections.

TABLE 124-2

Treatment of Common Infections of the Skin

DIAGNOSIS/CONDITION	PRIMARY TREATMENT	ALTERNATIVE TREATMENT	SEE ALSO CHAP(S).
Animal bite (prophylaxis or early infection) ^a	Amoxicillin–clavulanate (875/125 mg PO bid)	Doxycycline (100 mg PO bid)	136
Animal bite ^a (established infection)	Ampicillin–sulbactam (1.5–3 g IV q6h)	Clindamycin (600–900 mg IV q8h) <i>plus</i> Ciprofloxacin (400 mg IV q12h) <i>or</i> cefoxitin (2 g IV q6h)	136
Bacillary angiomatosis	Erythromycin (500 mg PO qid)	Doxycycline (100 mg PO bid)	167
Herpes simplex (primary genital)	Acyclovir (400 mg PO tid for 10 days)	Famciclovir (250 mg PO tid for 5–10 days) <i>or</i> valacyclovir (1000 mg PO bid for 10 days)	187
Herpes zoster (immunocompetent host >50 years of age)	Acyclovir (800 mg PO 5 times daily for 7–10 days)	Famciclovir (500 mg PO tid for 7–10 days) <i>or</i> valacyclovir (1000 mg PO tid for 7 days)	188
Cellulitis (staphylococcal or streptococcal ^{b,c})	Nafcillin or oxacillin (2 g IV q4–6h)	Cefazolin (1–2 g q8h) <i>or</i> ampicillin/sulbactam (1.5–3 g IV q6h) <i>or erythromycin</i> (0.5–1 g IV q6h) <i>or clindamycin</i> (600–900 mg IV q8h)	142, 143
MRSA skin infection ^d	Vancomycin (1 g IV q12h)	Linezolid (600 mg IV q12h)	142
Necrotizing fasciitis (group A streptococcal ^b)	Clindamycin (600–900 mg IV q6–8h) <i>plus</i> penicillin G (4 million units IV q4h)	Clindamycin (600–900 mg IV q6–8h) <i>plus</i> a cephalosporin (first- or second-generation)	143

DIAGNOSIS/CONDITION	PRIMARY TREATMENT	ALTERNATIVE TREATMENT	SEE ALSO CHAP(S).
Necrotizing fasciitis (mixed aerobes and anaerobes)	Ampicillin (2 g IV q4h) <i>plus</i> clindamycin (600–900 mg IV q6–8h) <i>plus</i> ciprofloxacin (400 mg IV q6–8h)	Vancomycin (1 g IV q6h) <i>plus</i> metronidazole (500 mg IV q6h) <i>plus</i> ciprofloxacin (400 mg IV q6–8h)	117 , 172
Gas gangrene	Clindamycin (600–900 mg IV q6–8h) <i>plus</i> penicillin G (4 million units IV q4–6h)	Clindamycin (600–900 mg IV q6–8h) <i>plus</i> cefoxitin (2 g IV q6h)	149

a

Pasteurella multocida, a species commonly associated with both dog and cat bites, is resistant to cephalexin, dicloxacillin, [clindamycin](#), and [erythromycin](#). *Eikenella corrodens*, a bacterium commonly associated with human bites, is resistant to [clindamycin](#), penicillinase-resistant penicillins, and [metronidazole](#) but is sensitive to trimethoprim-sulfamethoxazole and fluoroquinolones.

b

The frequency of [erythromycin](#) resistance in group A *Streptococcus* is currently ~5% in the United States but has reached 70–100% in some other countries. Most, but not all, erythromycin-resistant group A streptococci are susceptible to [clindamycin](#). Approximately 90% of *Staphylococcus aureus* strains are sensitive to [clindamycin](#), but resistance—both intrinsic and inducible—is increasing.

c

Severe hospital-acquired *S. aureus* infections or community-acquired *S. aureus* infections that are not responding to the β -lactam antibiotics recommended in this table may be caused by methicillin-resistant strains, requiring a switch to vancomycin, daptomycin, or linezolid.

d

Some strains of methicillin-resistant *S. aureus* (MRSA) remain sensitive to tetracycline and trimethoprim-sulfamethoxazole. Daptomycin (4 mg/kg IV q24h) or tigecycline (100-mg loading dose followed by 50 mg IV q12h) is an alternative treatment for MRSA.

FURTHER READING

Aldape MJ et al: *Clostridium sordellii* infection: Epidemiology, clinical findings, and current perspectives on diagnosis and treatment. Clin Infect Dis 43:1436, 2006. [[PubMed: 17083018](#)]

Stevens DL, Bryant AE: Life threatening skin and soft tissue infections, in *Netter's Infectious Diseases*, EC Jong, and DL Stevens (eds). Philadelphia, Elsevier, 2011, pp 94–101.

Stevens DL et al: Practice guidelines for the diagnosis and management of skin and soft tissue infections: 2014 update by the Infectious Diseases Society of America. Clin Infect Dis 59:e10, 2014 (Erratum: Clin Infect Dis 60:1448, 2015). [[PubMed: 24973422](#)]

Stevens DL et al: Trimethoprim-sulfamethoxazole vs placebo for the treatment of uncomplicated skin abscesses. N Engl J Med 374:824, 2016.

Talan DA et al: Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85, 1999. [[PubMed: 9887159](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 125: Infectious Arthritis

Lawrence C. Madoff

INTRODUCTION

Although *Staphylococcus aureus*, *Neisseria gonorrhoeae*, and other bacteria are the most common causes of infectious arthritis, various mycobacteria, spirochetes, fungi, and viruses also infect joints (**Table 125-1**). Since acute bacterial infection can destroy articular cartilage rapidly, all inflamed joints must be evaluated without delay to exclude noninfectious processes and determine appropriate antimicrobial therapy and drainage procedures. For more detailed information on infectious arthritis caused by specific organisms, the reader is referred to the chapters on those organisms.

TABLE 125-1

Differential Diagnosis of Arthritis Syndromes

ACUTE MONARTICULAR ARTHRITIS	CHRONIC MONARTICULAR ARTHRITIS	POLYARTICULAR ARTHRITIS
<i>Staphylococcus aureus</i> <i>Streptococcus pneumoniae</i> β -Hemolytic streptococci Gram-negative bacilli <i>Neisseria gonorrhoeae</i> <i>Candida</i> spp. Crystal-induced arthritis Fracture Hemarthrosis Foreign body Osteoarthritis Ischemic necrosis Monarticular rheumatoid arthritis	<i>Mycobacterium tuberculosis</i> Nontuberculous mycobacteria <i>Borrelia burgdorferi</i> <i>Treponema pallidum</i> <i>Candida</i> spp. <i>Sporothrix schenckii</i> <i>Coccidioides immitis</i> <i>Blastomyces dermatitidis</i> <i>Aspergillus</i> spp. <i>Cryptococcus neoformans</i> <i>Nocardia</i> spp. <i>Brucella</i> spp. Legg-Calvé-Perthes disease Osteoarthritis	<i>Neisseria meningitidis</i> <i>N. gonorrhoeae</i> Nongonococcal bacterial arthritis Bacterial endocarditis <i>Candida</i> spp. Poncet's disease (tuberculous rheumatism) Hepatitis B virus Parvovirus B19 HIV Human T-lymphotropic virus type 1 Rubella virus Arthropod-borne viruses Sickle cell disease flare Reactive arthritis Serum sickness Acute rheumatic fever Inflammatory bowel disease Systemic lupus erythematosus Rheumatoid arthritis/Still's disease Other vasculitides Sarcoidosis

Acute bacterial infection typically involves a single joint or a few joints. Subacute or chronic monarthritis or oligoarthritis suggests mycobacterial or fungal infection; episodic inflammation is seen in syphilis, Lyme disease, and the reactive arthritis that follows enteric infections and chlamydial urethritis. Acute polyarticular inflammation occurs as an immunologic reaction during the course of endocarditis, rheumatic fever, disseminated neisserial infection, and acute hepatitis B. Bacteria and viruses occasionally infect multiple joints, the former most commonly in persons with rheumatoid arthritis.

APPROACH TO THE PATIENT

APPROACH TO THE PATIENT**Infectious Arthritis**

Aspiration of synovial fluid—an essential element in the evaluation of potentially infected joints—can be performed without difficulty in most cases by the insertion of a large-bore needle into the site of maximal fluctuance or tenderness or by the route of easiest access. Ultrasonography or fluoroscopy may be used to guide aspiration of difficult-to-localize effusions of the hip and, occasionally, the shoulder and other joints. Normal synovial fluid contains <180 cells (predominantly mononuclear cells) per microliter. Synovial cell counts averaging 100,000/μL (range, 25,000–250,000/μL), with >90% neutrophils, are characteristic of acute bacterial infections. Crystal-induced, rheumatoid, and other noninfectious inflammatory arthritides usually are associated with <30,000–50,000 cells/μL; cell counts of 10,000–30,000/μL, with 50–70% neutrophils and the remainder lymphocytes, are common in mycobacterial and fungal infections. Definitive diagnosis of an infectious process relies on identification of the pathogen in stained smears of synovial fluid, isolation of the pathogen from cultures of synovial fluid and blood, or detection of microbial nucleic acids and proteins by nucleic acid amplification (NAA)–based assays and immunologic techniques.

ACUTE BACTERIAL ARTHRITIS**PATHOGENESIS**

Bacteria enter the joint from the bloodstream; from a contiguous site of infection in bone or soft tissue; or by direct inoculation during surgery, injection, animal or human bite, or trauma. In hematogenous infection, bacteria escape from synovial capillaries, which have no limiting basement membrane, and within hours provoke neutrophilic infiltration of the synovium. Neutrophils and bacteria enter the joint space; later, bacteria adhere to articular cartilage. Degradation of cartilage begins within 48 h as a result of increased intraarticular pressure, release of proteases and cytokines from chondrocytes and synovial macrophages, and invasion of the cartilage by bacteria and inflammatory cells. Histologic studies reveal bacteria lining the synovium and cartilage as well as abscesses extending into the synovium, cartilage, and—in severe cases—subchondral bone. Synovial proliferation results in the formation of a pannus over the cartilage, and thrombosis of inflamed synovial vessels develops. Bacterial factors that appear important in the pathogenesis of infective arthritis include various surface-associated adhesins in *S. aureus* that permit adherence to cartilage and endotoxins that promote chondrocyte-mediated breakdown of cartilage.

MICROBIOLOGY

The hematogenous route of infection is the most common route in all age groups, and nearly every bacterial pathogen is capable of causing septic arthritis. In infants, group B streptococci, gram-negative enteric bacilli, and *S. aureus* are the most common pathogens. Since the advent of the *Haemophilus influenzae* vaccine, the predominant causes among children <5 years of age have been *S. aureus*, *Streptococcus pyogenes* (group A *Streptococcus*), and (in some centers) *Kingella kingae*. Among young adults and adolescents, *N. gonorrhoeae* is the most commonly implicated organism. *S. aureus* accounts for most nongonococcal isolates in adults of all ages; gram-negative bacilli, pneumococci, and β-hemolytic streptococci—particularly groups A and B but also groups C, G, and F—are involved in up to one-third of cases in older adults, especially those with underlying comorbid illnesses.

Infections after surgical procedures or penetrating injuries are due most often to *S. aureus* and occasionally to other gram-positive bacteria or gram-negative bacilli. Infections with coagulase-negative staphylococci are unusual except after the implantation of prosthetic joints or arthroscopy. Anaerobic organisms, often in association with aerobic or facultative bacteria, are found after human bites and when decubitus ulcers or intraabdominal abscesses spread into adjacent joints. Polymicrobial infections complicate traumatic injuries with extensive contamination. Bites and scratches from cats and other animals may introduce *Pasteurella multocida* or *Bartonella henselae* into joints either directly or hematogenously, and bites from humans may introduce *Eikenella corrodens* or other components of the oral flora. Penetration of a sharp object through a shoe is associated with *Pseudomonas aeruginosa* arthritis in the foot.

NONGONOCOCCAL BACTERIAL ARTHRITIS**Epidemiology**

Although hematogenous infections with virulent organisms such as *S. aureus*, *H. influenzae*, and pyogenic streptococci occur in healthy persons, there is an underlying host predisposition in many cases of septic arthritis. Patients with rheumatoid arthritis have the highest incidence of infective arthritis (most often secondary to *S. aureus*) because of chronically inflamed joints; glucocorticoid therapy; and frequent breakdown of rheumatoid nodules, vasculitic ulcers, and skin overlying deformed joints. Diabetes mellitus, glucocorticoid therapy, hemodialysis, and malignancy all carry an increased risk of infection with *S. aureus* and gram-negative bacilli. Tumor necrosis factor inhibitors (e.g., etanercept, infliximab), which increasingly are used for the treatment of rheumatoid arthritis, predispose to mycobacterial infections and

possibly to other pyogenic bacterial infections and could be associated with septic arthritis in this population. Pneumococcal infections complicate alcoholism, deficiencies of humoral immunity, and hemoglobinopathies. Pneumococci, *Salmonella* species, and *H. influenzae* cause septic arthritis in persons infected with HIV. Persons with primary immunoglobulin deficiency are at risk for mycoplasmal arthritis, which results in permanent joint damage if tetracycline and replacement therapy with IV immunoglobulin are not administered promptly. IV drug users acquire staphylococcal and streptococcal infections from their own flora and acquire pseudomonal and other gram-negative infections from drugs and injection paraphernalia.

Clinical Manifestations

Some 90% of patients present with involvement of a single joint—most commonly the knee; less frequently the hip; and still less often the shoulder, wrist, or elbow. Small joints of the hands and feet are more likely to be affected after direct inoculation or a bite. Among IV drug users, infections of the spine, sacroiliac joints, and sternoclavicular joints (**Fig. 125-1**) are more common than infections of the appendicular skeleton. Polyarticular infection is most common among patients with rheumatoid arthritis and may resemble a flare of the underlying disease.

FIGURE 125-1

Acute septic arthritis of the sternoclavicular joint. A man in his forties with a history of cirrhosis presented with a new onset of fever and lower neck pain. He had no history of IV drug use or previous catheter placement. Jaundice and a painful swollen area over his left sternoclavicular joint were evident on physical examination. Cultures of blood drawn at admission grew group B *Streptococcus*. The patient recovered after treatment with IV penicillin. (Courtesy of Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The usual presentation consists of moderate to severe pain that is uniform around the joint, effusion, muscle spasm, and decreased range of motion. Fever in the range of 38.3–38.9°C (101–102°F) and sometimes higher is common but may not be present, especially in persons with rheumatoid arthritis, renal or hepatic insufficiency, or conditions requiring immunosuppressive therapy. The inflamed, swollen joint is usually evident on examination except in the case of a deeply situated joint such as the hip, shoulder, or sacroiliac joint. Cellulitis, bursitis, and acute osteomyelitis, which may produce a similar clinical picture, should be distinguished from septic arthritis by their greater range of motion and less-than-circumferential swelling. A focus of extraarticular infection, such as a boil or pneumonia, should be sought. Peripheral-blood leukocytosis with a left shift and elevation of the erythrocyte sedimentation rate or C-reactive protein level are common.

Plain radiographs show evidence of soft-tissue swelling, joint-space widening, and displacement of tissue planes by the distended capsule. Narrowing of the joint space and bony erosions indicate advanced infection and a poor prognosis. Ultrasound is useful for detecting effusions in the hip, and CT or MRI can demonstrate infections of the sacroiliac joint, the sternoclavicular joint, and the spine very well.

Laboratory Findings

Specimens of peripheral blood and synovial fluid should be obtained before antibiotics are administered. Blood cultures are positive in up to 50–70% of *S. aureus* infections but are less frequently positive in infections due to other organisms. The synovial fluid is turbid, serosanguineous, or frankly purulent. Gram-stained smears confirm the presence of large numbers of neutrophils. Levels of total protein and lactate dehydrogenase in synovial fluid are elevated, and the glucose level is depressed; however, these findings are not specific for infection, and measurement of these levels is not necessary for diagnosis. The synovial fluid should be examined for crystals because gout and

pseudogout can resemble septic arthritis clinically and infection and crystal-induced disease occasionally occur together. Organisms are seen on synovial fluid smears in nearly three-quarters of infections with *S. aureus* and streptococci and in 30–50% of infections due to gram-negative and other bacteria. Cultures of synovial fluid are positive in >90% of cases. Inoculation of synovial fluid into bottles containing liquid media for blood cultures increases the yield of a culture, especially if the pathogen is a fastidious organism or the patient is taking an antibiotic. NAA-based assays for bacterial DNA, when available, can be useful for the diagnosis of partially treated or culture-negative bacterial arthritis.

TREATMENT

TREATMENT

Nongonococcal Bacterial Arthritis

Prompt administration of systemic antibiotics and drainage of the involved joint can prevent destruction of cartilage, postinfectious degenerative arthritis, joint instability, or deformity. Once samples of blood and synovial fluid have been obtained for culture, empirical antibiotics should be directed against the bacteria visualized on smears or the pathogens that are likely in light of the patient's age and risk factors. Initial therapy should consist of IV-administered bactericidal agents; direct instillation of antibiotics into the joint is not necessary to achieve adequate levels in synovial fluid and tissue. An IV third-generation cephalosporin such as cefotaxime (1 g every 8 h) or ceftriaxone (1–2 g every 24 h) provides adequate empirical coverage for most community-acquired infections in adults when smears show no organisms. IV vancomycin (1 g every 12 h) is used if there are gram-positive cocci on the smear. If methicillin-resistant *S. aureus* is an unlikely pathogen (e.g., when it is not widespread in the community), cefazolin (2 g every 8 h), oxacillin (2 g every 4 h), or nafcillin (2 g every 4 h) should be given. In addition, an aminoglycoside or third-generation cephalosporin should be given to IV drug users and to other patients in whom *P. aeruginosa* may be the responsible agent.

Definitive therapy is based on the identity and antibiotic susceptibility of the bacteria isolated in culture. Infections due to staphylococci are treated with cefazolin, oxacillin, nafcillin, or vancomycin for 4 weeks. Pneumococcal and streptococcal infections due to penicillin-susceptible organisms respond to 2 weeks of therapy with penicillin G (2 million units IV every 4 h); infections caused by *H. influenzae* and by strains of *Streptococcus pneumoniae* that are resistant to penicillin are treated with cefotaxime or ceftriaxone for 2 weeks. Most enteric gram-negative infections can be cured in 3–4 weeks by a second- or third-generation cephalosporin given IV or by a fluoroquinolone such as [levofloxacin](#) (500 mg IV or PO every 24 h). *P. aeruginosa* infection should be treated for at least 2 weeks with a combination regimen composed of an aminoglycoside plus either an extended-spectrum penicillin such as mezlocillin (3 g IV every 4 h) or an antipseudomonal cephalosporin such as ceftazidime (1 g IV every 8 h). If tolerated, this regimen is continued for an additional 2 weeks; alternatively, a fluoroquinolone such as [ciprofloxacin](#) (750 mg PO twice daily) is given by itself or with the penicillin or cephalosporin in place of the aminoglycoside.

Timely drainage of pus and necrotic debris from the infected joint is required for a favorable outcome. Needle aspiration of readily accessible joints such as the knee may be adequate if loculations or particulate matter in the joint does not prevent its thorough decompression. Arthroscopic drainage and lavage may be employed initially or within several days if repeated needle aspiration fails to relieve symptoms, decrease the volume of the effusion and the synovial white cell count, and clear bacteria from smears and cultures. In some cases, arthrotomy is necessary to remove loculations and debride infected synovium, cartilage, or bone. Septic arthritis of the hip is best managed with arthrotomy, particularly in young children, in whom infection threatens the viability of the femoral head. Septic joints do not require immobilization except for pain control before symptoms are alleviated by treatment. Weight bearing should be avoided until signs of inflammation have subsided, but frequent passive motion of the joint is indicated to maintain full mobility. Although addition of glucocorticoids to antibiotic treatment improves the outcome of *S. aureus* arthritis in experimental animals, no clinical trials have evaluated this approach in humans.

GONOCOCCAL ARTHRITIS

Epidemiology

Although its incidence has declined in recent years, gonococcal arthritis ([Chap. 151](#)) has accounted for up to 70% of episodes of infectious arthritis in persons <40 years of age in the United States. Arthritis due to *N. gonorrhoeae* is a consequence of bacteremia arising from gonococcal infection or, more frequently, from asymptomatic gonococcal mucosal colonization of the urethra, cervix, or pharynx. Women are at greatest risk during menses and during pregnancy and overall are two to three times more likely than men to develop disseminated gonococcal infection (DGI) and arthritis. Persons with complement deficiencies, especially of the terminal components, are prone to recurrent episodes of gonococcemia. Strains of gonococci that are most likely to cause DGI include those which produce transparent colonies in culture, have the type IA outer-membrane protein, or are of the AUH-auxotroph type.

Clinical Manifestations and Laboratory Findings

The most common manifestation of DGI is a syndrome of fever, chills, rash, and articular symptoms. Small numbers of papules that progress to hemorrhagic pustules develop on the trunk and the extensor surfaces of the distal extremities. Migratory arthritis and tenosynovitis of the knees, hands, wrists, feet, and ankles are prominent. The cutaneous lesions and articular findings are believed to be the consequence of an immune reaction to circulating gonococci and immune-complex deposition in tissues. Thus, cultures of synovial fluid are consistently negative, and blood cultures are positive in fewer than 45% of patients. Synovial fluid may be difficult to obtain from inflamed joints and usually contains only 10,000–20,000 leukocytes/ μ L.

True gonococcal septic arthritis is less common than the DGI syndrome and always follows DGI, which is unrecognized in one-third of patients. A single joint such as the hip, knee, ankle, or wrist is usually involved. Synovial fluid, which contains $>50,000$ leukocytes/ μ L, can be obtained with ease; the gonococcus is evident only occasionally in Gram-stained smears, and cultures of synovial fluid are positive in fewer than 40% of cases. Blood cultures are almost always negative.

Because it is difficult to isolate gonococci from synovial fluid and blood, specimens for culture should be obtained from potentially infected mucosal sites. NAA-based urine tests also may be positive. Cultures and Gram-stained smears of skin lesions are occasionally positive. All specimens for culture should be plated onto Thayer-Martin agar directly or in special transport media at the bedside and transferred promptly to the microbiology laboratory in an atmosphere of 5% CO₂, as generated in a candle jar. NAA-based assays are extremely sensitive in detecting gonococcal DNA in synovial fluid. A dramatic alleviation of symptoms within 12–24 h after the initiation of appropriate antibiotic therapy supports a clinical diagnosis of the DGI syndrome if cultures are negative.

TREATMENT

TREATMENT

Gonococcal Arthritis

Initial treatment consists of ceftriaxone (1 g IV or IM every 24 h) to cover possible penicillin-resistant organisms. Once local and systemic signs are clearly resolving, the 7-day course of therapy can be completed with an oral fluoroquinolone such as [ciprofloxacin](#) (500 mg twice daily) if the organism is known to be susceptible. If penicillin-susceptible organisms are isolated, amoxicillin (500 mg three times daily) may be used. Suppurative arthritis usually responds to needle aspiration of involved joints and 7–14 days of antibiotic treatment. Arthroscopic lavage or arthrotomy is rarely required. Patients with DGI should be treated for *Chlamydia trachomatis* infection unless this infection is ruled out by appropriate testing. Addition of [azithromycin](#) (1 g orally as a single dose) is recommended to treat chlamydial co-infection, which is common. Sexual partners should be offered testing and presumptive treatment for gonorrhea and chlamydial infection.

It is noteworthy that arthritis symptoms similar to those seen in DGI occur in meningococcemia. A dermatitis–arthritis syndrome, purulent monoarthritis, and reactive polyarthritis have been described. All respond to treatment with IV penicillin.

SPIROCHETAL ARTHRITIS

LYME DISEASE

Lyme disease ([Chap. 181](#)) due to infection with the spirochete *Borrelia burgdorferi* causes arthritis in up to 60% of persons who are not treated. Intermittent arthralgias and myalgias—but not arthritis—occur within days or weeks of inoculation of the spirochete by the *Ixodes* tick. Later, there are three patterns of joint disease: (1) Fifty percent of untreated persons experience intermittent episodes of monoarthritis or oligoarthritis involving the knee and/or other large joints. The symptoms wax and wane without treatment over months, and each year 10–20% of patients report loss of joint symptoms. (2) Twenty percent of untreated persons develop a pattern of waxing and waning arthralgias. (3) Ten percent of untreated patients develop chronic inflammatory synovitis that results in erosive lesions and destruction of the joint. Serologic tests for IgG antibodies to *B. burgdorferi* are positive in more than 90% of persons with Lyme arthritis, and an NAA-based assay detects *Borrelia* DNA in 85%.

TREATMENT

TREATMENT

Lyme Arthritis

Lyme arthritis generally responds well to therapy. A regimen of oral doxycycline (100 mg twice daily for 28 days), oral amoxicillin (500 mg three times daily for 28 days), or parenteral ceftriaxone (2 g/d for 2–4 weeks) is recommended. Patients who do not respond to a total of 2 months of oral therapy or 1 month of parenteral therapy are unlikely to benefit from additional antibiotic therapy and are treated with anti-inflammatory

agents or synovectomy. Failure of therapy is associated with host features such as the human leukocyte antigen DR4 (HLA-DR4) genotype, persistent reactivity to OspA (outer-surface protein A), and the presence of hLFA-1 (human leukocyte function-associated antigen 1), which cross-reacts with OspA.

SYPHILITIC ARTHRITIS

Articular manifestations occur in different stages of syphilis (**Chap. 177**). In early congenital syphilis, periarticular swelling and immobilization of the involved limbs (*Parrot's pseudoparalysis*) complicate osteochondritis of long bones. *Clutton's joint*, a late manifestation of congenital syphilis that typically develops between ages 8 and 15 years, is caused by chronic painless synovitis with effusions of large joints, particularly the knees and elbows. Secondary syphilis may be associated with arthralgias, with symmetric arthritis of the knees and ankles and occasionally of the shoulders and wrists, and with sacroiliitis. The arthritis follows a subacute to chronic course with a mixed mononuclear and neutrophilic synovial-fluid pleocytosis (typical cell counts, 5000–15,000/ μ L). Immunologic mechanisms may contribute to the arthritis, and symptoms usually improve rapidly with penicillin therapy. In tertiary syphilis, Charcot joint results from sensory loss due to tabes dorsalis. Penicillin is not helpful in this setting.

MYCOBACTERIAL ARTHRITIS

Tuberculous arthritis (**Chap. 173**) accounts for ~1% of all cases of tuberculosis and 10% of extrapulmonary cases. The most common presentation is chronic granulomatous monoarthritis. An unusual syndrome, *Poncet's disease*, is a reactive symmetric form of polyarthritis that affects persons with visceral or disseminated tuberculosis. No mycobacteria are found in the joints, and symptoms resolve with antituberculous therapy.

Unlike tuberculous osteomyelitis (**Chap. 126**), which typically involves the thoracic and lumbar spine (50% of cases), tuberculous arthritis primarily involves the large weight-bearing joints, in particular the hips, knees, and ankles, and only occasionally involves smaller non-weight-bearing joints. Progressive monoarticular swelling and pain develop over months or years, and systemic symptoms are seen in only half of all cases. Tuberculous arthritis occurs as part of a disseminated primary infection or through late reactivation, often in persons with HIV infection or other immunocompromised hosts. Coexistent active pulmonary tuberculosis is unusual.

Aspiration of the involved joint yields fluid with an average cell count of 20,000/ μ L, with ~50% neutrophils. Acid-fast staining of the fluid yields positive results in fewer than one-third of cases, and cultures are positive in 80%. Culture of synovial tissue taken at biopsy is positive in ~90% of cases and shows granulomatous inflammation in most. NAA methods can shorten the time to diagnosis to 1 or 2 days. Radiographs reveal peripheral erosions at the points of synovial attachment, periarticular osteopenia, and eventually joint-space narrowing. Therapy for tuberculous arthritis is the same as that for tuberculous pulmonary disease, requiring the administration of multiple agents for 6–9 months. Therapy is more prolonged in immunosuppressed individuals, such as those infected with HIV.

Various atypical mycobacteria (**Chap. 175**) found in water and soil may cause chronic indolent arthritis. Such disease results from trauma and direct inoculation associated with farming, gardening, or aquatic activities. Smaller joints, such as the digits, wrists, and knees, are usually involved. Involvement of tendon sheaths and bursae is typical. The mycobacterial species involved include *Mycobacterium marinum*, *M. avium-intracellulare*, *M. terrae*, *M. kansasii*, *M. fortuitum*, and *M. chelonae*. In persons who have HIV infection or are receiving immunosuppressive therapy, hematogenous spread to the joints has been reported for *M. kansasii*, *M. avium* complex, and *M. haemophilum*. Diagnosis usually requires biopsy and culture, and therapy is based on antimicrobial susceptibility patterns.

FUNGAL ARTHRITIS

Fungi are an unusual cause of chronic monoarticular arthritis. Granulomatous articular infection with the endemic dimorphic fungi *Coccidioides immitis*, *Blastomyces dermatitidis*, and (less commonly) *Histoplasma capsulatum* (**Fig. 125-2**) results from hematogenous seeding or direct extension from bony lesions in persons with disseminated disease. Joint involvement is an unusual complication of sporotrichosis (infection with *Sporothrix schenckii*) among gardeners and other persons who work with soil or sphagnum moss. Articular sporotrichosis is six times more common among men than among women, and alcoholics and other debilitated hosts are at risk for polyarticular infection.

FIGURE 125-2

Chronic arthritis caused by *Histoplasma capsulatum* in the left knee. A. A man in his sixties from El Salvador presented with a history of progressive knee pain and difficulty walking for several years. He had undergone arthroscopy for a meniscal tear 7 years before presentation (without relief) and had received several intraarticular glucocorticoid injections. The patient developed significant deformity of the knee over time, including a large effusion in the lateral aspect. **B.** An x-ray of the knee showed multiple abnormalities, including severe medial

femorotibial joint-space narrowing, several large subchondral cysts within the tibia and the patellofemoral compartment, a large suprapatellar joint effusion, and a large soft-tissue mass projecting laterally over the knee. **C.** MRI further defined these abnormalities and demonstrated the cystic nature of the lateral knee abnormality. Synovial biopsies demonstrated chronic inflammation with giant cells, and cultures grew *H. capsulatum* after 3 weeks of incubation. All clinical cystic lesions and the effusion resolved after 1 year of treatment with itraconazole. The patient underwent a left total-knee replacement for definitive treatment. (Courtesy of Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)



A Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Candida infection involving a single joint—usually the knee, hip, or shoulder—results from surgical procedures, intraarticular injections, or (among critically ill patients with debilitating illnesses such as diabetes mellitus or hepatic or renal insufficiency and patients receiving immunosuppressive therapy) hematogenous spread. *Candida* infections in IV drug users typically involve the spine, sacroiliac joints, or other fibrocartilaginous joints. Unusual cases of arthritis due to *Aspergillus* species, *Cryptococcus neoformans*, *Pseudallescheria boydii*, and the dematiaceous fungi also have resulted from direct inoculation or disseminated hematogenous infection in immunocompromised persons. In the United States, a 2012 national outbreak of fungal arthritis (and meningitis) caused by *Exserohilum rostratum* was linked to intraspinal and intraarticular injection of a contaminated preparation of methylprednisolone acetate.

The synovial fluid in fungal arthritis usually contains 10,000–40,000 cells/ μ L, with ~70% neutrophils. Stained specimens and cultures of synovial tissue often confirm the diagnosis of fungal arthritis when studies of synovial fluid give negative results. Treatment consists of drainage and lavage of the joint and systemic administration of an antifungal agent directed at a specific pathogen. The doses and duration of therapy are the same as for disseminated disease (see Part 5, Section 16). Intraarticular instillation of amphotericin B has been used in addition to IV therapy.

VIRAL ARTHRITIS

Viruses produce arthritis by infecting synovial tissue during systemic infection or by provoking an immunologic reaction that involves joints. As many as 50% of women report persistent arthralgias and 10% report frank arthritis within 3 days of the rash that follows natural infection with rubella virus and within 2–6 weeks after receipt of live-virus vaccine. Episodes of symmetric inflammation of fingers, wrists, and knees uncommonly recur for >1 year, but a syndrome of chronic fatigue, low-grade fever, headaches, and myalgias can persist for months or years. IV immunoglobulin has been helpful in selected cases. Self-limited monarticular or migratory polyarthritis may develop within 2 weeks of the parotitis of mumps; this sequela is more common among men than among women. Approximately 10% of children and 60% of women develop arthritis after infection with parvovirus B19. In adults, arthropathy sometimes occurs without fever or rash. Pain and stiffness, with less prominent swelling (primarily of the hands but also of the knees, wrists, and ankles), usually resolve within weeks, although a small proportion of patients develop chronic arthropathy.

About 2 weeks before the onset of jaundice, up to 10% of persons with acute hepatitis B develop an immune complex-mediated, serum sickness-like reaction with maculopapular rash, urticaria, fever, and arthralgias. Less common developments include symmetric arthritis involving the hands, wrists, elbows, or ankles and morning stiffness that resembles a flare of rheumatoid arthritis. Symptoms resolve at the time jaundice develops. Many persons with chronic hepatitis C infection report persistent arthralgia or arthritis, both in the presence and in the absence of cryoglobulinemia.



Painful arthritis involving larger joints often accompanies the fever and rash of several arthropod-borne viral infections, including those caused by Zika, chikungunya, O'nyong-nyong, Ross River, Mayaro, and Barmah Forest viruses (Chap. 204). Symmetric arthritis involving the

hands and wrists may occur during the convalescent phase of infection with lymphocytic choriomeningitis virus. Patients infected with an enterovirus frequently report arthralgias, and echovirus has been isolated from patients with acute polyarthrititis.

Several arthritis syndromes are associated with HIV infection. Reactive arthritis with painful lower-extremity oligoarthritis often follows an episode of urethritis in HIV-infected persons. HIV-associated reactive arthritis appears to be extremely common among persons with the HLA-B27 haplotype, but sacroiliac joint disease is unusual and is seen mostly in the absence of HLA-B27. Up to one-third of HIV-infected persons with psoriasis develop psoriatic arthritis. Painless monoarthropathy and persistent symmetric polyarthropathy occasionally complicate HIV infection. Chronic persistent oligoarthritis of the shoulders, wrists, hands, and knees occurs in women infected with human T-lymphotropic virus type 1. Synovial thickening, destruction of articular cartilage, and leukemic-appearing atypical lymphocytes in synovial fluid are characteristic, but progression to T cell leukemia is unusual.

PARASITIC ARTHRITIS

Arthritis due to parasitic infection is rare. The guinea worm *Dracunculus medinensis* may cause destructive joint lesions in the lower extremities as migrating gravid female worms invade joints or cause ulcers in adjacent soft tissues that become secondarily infected. Hydatid cysts infect bones in 1–2% of cases of infection with *Echinococcus granulosus*. The expanding destructive cystic lesions may spread to and destroy adjacent joints, particularly the hip and pelvis. In rare cases, chronic synovitis has been associated with the presence of schistosomal eggs in synovial biopsies. Monoarticular arthritis in children with lymphatic filariasis appears to respond to therapy with diethylcarbamazine even in the absence of microfilariae in synovial fluid. Reactive arthritis has been attributed to hookworm, *Strongyloides*, *Cryptosporidium*, and *Giardia* infection in case reports, but confirmation is required.

POSTINFECTIOUS OR REACTIVE ARTHRITIS

Reactive polyarthrititis develops several weeks after ~1% of cases of nongonococcal urethritis and 2% of enteric infections, particularly those due to *Yersinia enterocolitica*, *Shigella flexneri*, *Campylobacter jejuni*, and *Salmonella* species. Only a minority of these patients have the other findings of classic reactive arthritis, including urethritis, conjunctivitis, uveitis, oral ulcers, and rash. Studies have identified microbial DNA or antigen in synovial fluid or blood, but the pathogenesis of this condition is poorly understood.

Reactive arthritis is most common among young men (except after *Yersinia* infection) and has been linked to the HLA-B27 locus as a potential genetic predisposing factor. Patients report painful, asymmetric oligoarthritis that affects mainly the knees, ankles, and feet. Low-back pain is common, and radiographic evidence of sacroiliitis is found in patients with long-standing disease. Most patients recover within 6 months, but prolonged recurrent disease is more common in cases that follow chlamydial urethritis. Anti-inflammatory agents help relieve symptoms, but the role of prolonged antibiotic therapy in eliminating microbial antigen from the synovium is controversial.

Migratory polyarthrititis and fever constitute the usual presentation of acute rheumatic fever in adults (Chap. 352). This presentation is distinct from that of poststreptococcal reactive arthritis, which also follows infections with group A *Streptococcus* but is not migratory, lasts beyond the typical 3-week maximum of acute rheumatic fever, and responds poorly to aspirin.

INFECTIONS IN PROSTHETIC JOINTS

Infection complicates 1–4% of total joint replacements. The majority of infections are acquired intraoperatively or immediately postoperatively as a result of wound breakdown or infection; less commonly, these joint infections develop later after joint replacement and are the result of hematogenous spread or direct inoculation. The presentation may be acute, with fever, pain, and local signs of inflammation, especially in infections due to *S. aureus*, pyogenic streptococci, and enteric bacilli. Alternatively, infection may persist for months or years without causing constitutional symptoms when less virulent organisms, such as coagulase-negative staphylococci or diphtheroids, are involved. Such indolent infections usually are acquired during joint implantation and are discovered during evaluation of chronic unexplained pain or after a radiograph shows loosening of the prosthesis; the erythrocyte sedimentation rate and C-reactive protein level are usually elevated in such cases.

The diagnosis is best made by needle aspiration of the joint; accidental introduction of organisms during aspiration must be avoided meticulously. Synovial fluid pleocytosis with a predominance of polymorphonuclear leukocytes is highly suggestive of infection, since other inflammatory processes uncommonly affect prosthetic joints. Culture and Gram's stain usually yield the responsible pathogen. Sonication of explanted prosthetic material can improve the yield of culture, presumably by breaking up bacterial biofilms on the surfaces of prostheses. Use of special media for unusual pathogens such as fungi, atypical mycobacteria, and *Mycoplasma* may be necessary if routine and anaerobic cultures are negative.

TREATMENT**TREATMENT****Prosthetic Joint Infections**

Treatment includes surgery and high doses of parenteral antibiotics, which are given for 4–6 weeks because bone is usually involved. In most cases, the prosthesis must be replaced to cure the infection. Implantation of a new prosthesis is best delayed for several weeks or months because relapses of infection occur most commonly within this time frame. In some cases, reimplantation is not possible, and the patient must manage without a joint, with a fused joint, or even with amputation. Cure of infection without removal of the prosthesis is occasionally possible in cases that are due to streptococci or pneumococci and that lack radiologic evidence of loosening of the prosthesis. In these cases, antibiotic therapy must be initiated within several days of the onset of infection, and the joint should be drained vigorously by open arthrotomy or arthroscopically. In selected patients who prefer to avoid the high morbidity rate associated with joint removal and reimplantation, suppression of the infection with antibiotics may be a reasonable goal. A high cure rate with retention of the prosthesis has been reported when the combination of oral rifampin and another antibiotic (e.g., a quinolone, an antistaphylococcal penicillin, or vancomycin) is given for 3–6 months to persons with staphylococcal prosthetic joint infection of short duration. This approach, which is based on the ability of rifampin to kill organisms adherent to foreign material and in the stationary growth phase, requires confirmation in prospective trials.

PREVENTION

To avoid the disastrous consequences of infection, candidates for joint replacement should be selected with care. Rates of infection are particularly high among patients with rheumatoid arthritis, persons who have undergone previous surgery on the joint, and persons with medical conditions requiring immunosuppressive therapy. Perioperative antibiotic prophylaxis, usually with cefazolin, and measures to decrease intraoperative contamination, such as laminar flow, have lowered the rates of perioperative infection to <1% in many centers. After implantation, measures should be taken to prevent or rapidly treat extra-articular infections that might give rise to hematogenous spread to the prosthesis. The effectiveness of prophylactic antibiotics for the prevention of hematogenous infection after dental procedures has not been demonstrated; in fact, viridans streptococci and other components of the oral flora are extremely unusual causes of prosthetic joint infection. Accordingly, the American Dental Association and the American Academy of Orthopaedic Surgeons do not recommend antibiotic prophylaxis for most dental patients with total joint replacements and have stated that there is no convincing evidence to support its use. Similarly, guidelines issued by the American Urological Association and the American Academy of Orthopaedic Surgeons do not recommend the use of prophylactic antibiotics for most patients with prosthetic joints who are undergoing urologic procedures but state that prophylaxis should be considered in certain situations—e.g., for patients (especially immunocompromised patients) who are undergoing a procedure posing a relatively high risk of bacteremia (such as lithotripsy or surgery involving bowel segments).

Acknowledgments

The contributions of James H. Maguire and the late Scott J. Thaler to this chapter in earlier editions are gratefully acknowledged.

FURTHER READING

Bardin T: Gonococcal arthritis. Best Pract Res Clin Rheumatol 17:201, 2003. [PubMed: 12787521]

Borzio R et al: Predictors of septic arthritis in the adult population. Orthopedics 39:e657, 2016. [PubMed: 27286047]

Franssila R, Hedman K: Infection and musculoskeletal conditions: Viral causes of arthritis. Best Pract Res Clin Rheumatol 20:1139, 2006. [PubMed: 17127201]

Harrington JT: Mycobacterial and fungal arthritis. Curr Opin Rheumatol 10:335, 1998. [PubMed: 9725095]

Meehan AM et al: Outcome of penicillin-susceptible streptococcal prosthetic joint infection treated with debridement and retention of the prosthesis. Clin Infect Dis 36:845, 2003. [PubMed: 12652384]

Zeller V et al: One-stage exchange arthroplasty for chronic periprosthetic hip infection: Results of a large prospective cohort study. J Bone Joint Surg Am 96:e1, 2014. [PubMed: 24382729]

Zimmerli W et al: Prosthetic-joint infections. N Engl J Med 351:145, 2004. [PubMed: 15247354]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 126: Osteomyelitis

Werner Zimmerli

INTRODUCTION

Osteomyelitis, an infection of bone, can be caused by various microorganisms that arrive at bone through different routes. Spontaneous hematogenous osteomyelitis may occur in otherwise healthy individuals, whereas local microbial spread mainly affects either individuals who have underlying disease (e.g., vascular insufficiency) or patients who have compromised skin or other tissue barriers, with consequent exposure of bone. The latter situation typically follows surgery involving bone, such as sternotomy or orthopedic repair.

The manifestations of osteomyelitis are different in children and adults. In children circulating microorganisms seed mainly long bones, whereas in adults the vertebral column is the most commonly affected site.

Management of osteomyelitis differs greatly depending on whether an implant is involved. The most important aim of the management of either type of osteomyelitis is to prevent progression to chronic osteomyelitis by rapid diagnosis and prompt treatment. Device-related bone and joint infection necessitates a multidisciplinary approach requiring antibiotic therapy and, in many cases, surgical removal of the device. For most types of osteomyelitis, the optimal duration of antibiotic treatment has not been established in clinical trials. Therefore, the recommendations for therapy in this chapter reflect mainly expert opinions.

CLASSIFICATION

There is no generally accepted, comprehensive system for classification of osteomyelitis, primarily because of the multifaceted presentation of this infection. Different specialists are confronted with different facets of bone disease. Most often, however, general practitioners or internists are the first to encounter patients with the initial signs and symptoms of osteomyelitis. These primary care physicians should be able to recognize this disease in any of its forms. Osteomyelitis cases can be classified by various criteria, including pathogenesis, duration of infection, location of infection, and presence or absence of foreign material. The widely used Cierny-Mader staging system classifies osteomyelitis according to anatomic site, comorbidity, and radiographic findings, with stratification of long-bone osteomyelitis to optimize surgical management; this system encompasses both systemic and local factors affecting immune status, metabolism, and local vascularity.

Any of three mechanisms can underlie osteomyelitis: (1) hematogenous spread; (2) spread from a contiguous site following surgery; and (3) secondary infection in the setting of vascular insufficiency or concomitant neuropathy. Hematogenous osteomyelitis in adults typically involves the vertebral column. In only about half of patients a primary focus can be detected. The most common primary foci of infection are the urinary tract, skin/soft tissue, intravascular catheterization sites, and the endocardium. Spread from a contiguous source follows either bone trauma or surgical intervention. Wound infection leading to osteomyelitis typically occurs after cardiovascular intervention involving the sternum, orthopedic repair after open fracture, or prosthetic joint insertion. Osteomyelitis secondary to vascular insufficiency or peripheral neuropathy most often follows chronic, progressively deep skin and soft tissue infection of the foot. The most common underlying condition is diabetes. In diabetes that is poorly controlled, the *diabetic foot syndrome* is caused by skin, soft tissue, and bone ischemia combined with motor, sensory, and autonomic neuropathy.

Classification of osteomyelitis according to the duration of infection, although ill defined, is useful because the management of acute and chronic osteomyelitis differs. Whereas acute osteomyelitis can generally be treated with antibiotics alone, antibiotic treatment for chronic osteomyelitis should be combined with debridement surgery. Acute hematogenous or contiguous osteomyelitis evolves over a short period—i.e., a few days or weeks. In contrast, subacute or chronic osteomyelitis lasts for weeks or months before treatment is started. Typical examples of a subacute course are vertebral osteomyelitis due to tuberculosis or brucellosis and delayed implant-associated infections caused mainly by low-virulence microorganisms (coagulase-negative staphylococci, *Propionibacterium acnes*). Chronic osteomyelitis develops when insufficient therapy leads to persistence or recurrence, most often after sternal, mandibular, or foot infection.

Classification by location distinguishes among cases in the long bones, the vertebral column, and the periarticular bones. Long bones are generally involved after hematogenous seeding in children or contiguous spread following trauma or surgery. The risk of vertebral osteomyelitis in adults increases with age. Periarticular osteomyelitis, which complicates septic arthritis that has not been adequately treated, is especially common in periprosthetic joint infection.

Osteomyelitis involving a foreign device requires surgical management for cure. Even acute implant-associated infection calls for prolonged antimicrobial therapy. Therefore, identification of this type of disease is of practical importance.

VERTEBRAL OSTEOMYELITIS

PATHOGENESIS

Vertebral osteomyelitis, also referred to as disk-space infection, septic diskitis, spondylodiskitis, or spinal osteomyelitis, is the most common manifestation of hematogenous bone infection in adults. This designation reflects a pathogenic process leading to involvement of the adjacent vertebrae and the corresponding intervertebral disk. In adults, the disk is avascular. Microorganisms invade via the segmental arterial circulation in adjacent endplates and then spread into the disk. Alternative routes of infection are retrograde seeding through the prevertebral venous plexus and direct inoculation during spinal surgery,

epidural infiltration, or trauma. In the setting of implant surgery, microorganisms are inoculated either during the procedure or, if wound healing is impaired, in the early postoperative period.

EPIDEMIOLOGY

Vertebral osteomyelitis occurs more often in male than in female patients (ratio, 1.5:1). Between 1995 and 2008, the incidence rate increased from 2.2 to 5.8 cases/100,000 person-years. There is a clear age-dependent increase from 0.3 case/100,000 at ages <20 years to 6.5 cases/100,000 at ages >70 years. The observed increase in reported cases during the past two decades may reflect improvements in diagnosis resulting from the broad availability of MRI technology. In addition, the fraction of cases of vertebral osteomyelitis acquired in association with health care is increasing as a consequence of comorbidity and the rising number of invasive interventions.

MICROBIOLOGY



Vertebral osteomyelitis is typically classified as pyogenic or nonpyogenic. However, this distinction is arbitrary: in “nonpyogenic” cases (tuberculous, brucellar), macroscopic pus formation (caseous necrosis, abscess) is quite common. A more accurate scheme is to classify cases as acute or subacute/chronic. Whereas the microbiologic spectrum of acute cases is similar in different parts of the world, the spectrum of subacute/chronic cases varies according to the geographic region. The great majority of cases are monomicrobial in etiology. Of episodes of acute vertebral osteomyelitis, 40–50% are caused by *Staphylococcus aureus*, 12% by streptococci, and 20% by gram-negative bacilli—mainly *Escherichia coli* (9%) and *Pseudomonas aeruginosa* (6%). Subacute vertebral osteomyelitis is typically caused by *Mycobacterium tuberculosis* or *Brucella* species in regions where these microorganisms are endemic. Osteomyelitis due to viridans streptococci also has a subacute presentation; these infections most often occur as secondary foci in patients with endocarditis. In vertebral osteomyelitis due to *Candida* species, the diagnosis is often delayed by several weeks; this etiology should be suspected in IV drug users who do not use sterile paraphernalia. In implant-associated spinal osteomyelitis, coagulase-negative staphylococci and *P. acnes*—which, in the absence of an implant, are generally considered contaminants—typically cause low-grade (chronic) infections. As an exception, coagulase-negative staphylococci can cause native spinal osteomyelitis in cases of prolonged bacteremia (e.g., in patients with infected pacemaker electrodes or implanted vascular catheters that are not promptly removed).

CLINICAL MANIFESTATIONS

The signs and symptoms of vertebral osteomyelitis are nonspecific. Only about half of patients develop fever >38°C (>100.4°F), perhaps because patients frequently use analgesic drugs. Back pain is the leading initial symptom (>85% of cases). The location of the pain corresponds to the site of infection: the cervical spine in ~10% of cases, the thoracic spine in 30%, and the lumbar spine in 60%. One exception is involvement at the thoracic level in two-thirds of cases of tuberculous osteomyelitis and at the lumbar level in only one-third.

This difference is due to direct mycobacterial spread via pleural or mediastinal lymph nodes in pulmonary tuberculosis.

Neurologic deficits, such as radiculopathy, weakness, or sensory loss, are observed in about one-third of cases of vertebral osteomyelitis. In brucellar vertebral osteomyelitis, neurologic impairment is less common; in tuberculous osteomyelitis, it is about twice as common as in cases of other etiologies. Neurologic signs and symptoms are caused mostly by spinal epidural abscess. This complication starts with severe localized back pain and progresses to radicular pain, reflex changes, sensory abnormalities, motor weakness, bowel and bladder dysfunction, and paralysis.

A primary focus should always be sought but is found in only half of cases. Overall, endocarditis is identified in ~10% of patients. In osteomyelitis caused by viridans streptococci, endocarditis is the source in about half of patients.

Implant-associated spinal osteomyelitis can present as either early- or late-onset infection. Early-onset infection is diagnosed within 30 days after implant placement. *S. aureus* is the most common pathogen. Wound healing impairment and fever are the leading findings. Late-onset infection is diagnosed beyond 30 days after surgery, with low-virulence organisms such as coagulase-negative staphylococci or *P. acnes* as typical infecting agents. Fever is rare. One-quarter of patients have a sinus tract. Because of the delayed course and the lack of classic signs of infection, rapid diagnosis requires a high degree of suspicion.

DIAGNOSIS

Leukocytosis and neutrophilia have low levels of diagnostic sensitivity (only 65% and 40%, respectively). In contrast, an increased erythrocyte sedimentation rate or C-reactive protein (CRP) level has been reported in 98% and 100% of cases, respectively; thus, these tests are helpful in excluding vertebral osteomyelitis. The fraction of blood cultures that yield positive results depends heavily on whether the patient has been pretreated with antibiotics; across studies, the range is from 30% to 78%. In view of this low rate of positive blood culture after antibiotic treatment, such therapy should be withheld until microbial growth is proven unless the patient has sepsis syndrome. In patients with negative blood cultures, CT-guided or open biopsy is needed. Whether a CT-guided biopsy with a negative result is repeated or followed by open biopsy depends on the experience of personnel at the specific center. Bone samples should be cultured for aerobic, anaerobic, and fungal agents, with a portion of the sample sent for histopathologic study. In cases with a subacute/chronic presentation, a suggestive history, or a granuloma detected during histopathologic analysis, mycobacteria and brucellae also should be sought. When blood and tissue cultures are negative despite suggestive histopathology, broad-range polymerase chain reaction analysis of biopsy specimens or aspirated pus should be considered. This technique allows detection of unusual pathogens such as *Tropheryma whipplei*.

Given that signs and symptoms of osteomyelitis are nonspecific, the clinical differential diagnosis of febrile back pain is broad, including pyelonephritis, pancreatitis, and viral syndromes. In addition, multiple noninfectious pathologies of the vertebral column, such as osteoporotic fracture, seronegative spondylitis

(ankylosing spondylitis, psoriasis, reactive arthritis, enteropathic arthritis), and spinal stenosis must be considered.

Imaging procedures are the most important tools not only for the diagnosis of vertebral osteomyelitis but also for the detection of pyogenic complications and alternative conditions (e.g., bone metastases or osteoporotic fractures). Plain radiography is a reasonable first step in evaluating patients without neurologic symptoms and may reveal an alternative diagnosis. Because of its low sensitivity, plain radiography generally is not helpful in acute osteomyelitis, but it can be useful in subacute or chronic cases. The gold standard is MRI, which should be performed expeditiously in patients with neurologic impairment in order to rule out a herniated disk or to detect pyogenic complications in a timely manner. Even if the pathologic findings on MRI suggest vertebral osteomyelitis, alternative diagnoses should be considered, especially when blood cultures are negative. The most common alternative diagnosis is erosive osteochondrosis. Septic bone necrosis, gouty spondylodiskitis, and erosive diskovertebral lesions (Andersson lesions) in ankylosing spondylitis may likewise mimic vertebral osteomyelitis. CT is less sensitive than MRI but may be helpful in guiding a percutaneous biopsy. In the future, positron-emission tomography (PET) with ^{18}F -fluorodeoxyglucose, which has a high degree of diagnostic accuracy, may be an alternative imaging procedure when MRI is contraindicated. ^{18}F -fluorodeoxyglucose PET should be considered for patients with implants and patients in whom several foci are suspected.

TREATMENT

TREATMENT

Vertebral Osteomyelitis

The aims of therapy for vertebral osteomyelitis are (1) elimination of the pathogen(s), (2) protection from further bone loss, (3) relief of back pain, (4) prevention of complications, and (5) stabilization, if needed.

Table 126-1 summarizes suggested antimicrobial regimens for infections attributable to the most common etiologic agents. For optimal antimicrobial therapy, identification of the infecting agent is required. Therefore, in patients without sepsis syndrome, antibiotics should not be administered until the pathogen is identified in a blood culture, a bone biopsy, or an aspirated pus collection. Traditionally, bone infections are at least initially treated by the IV route. Unfortunately, relevant controlled trials are lacking, and the preference for the IV route is not evidence based. There are no good arguments for the assumption that IV therapy is superior to oral administration if the following requirements are met: (1) optimal antibiotic spectrum, (2) excellent bioavailability of the oral drug, (3) clinical studies confirming efficacy of the oral drug, (4) normal intestinal function, and (5) no vomiting. However, a short initial course of parenteral therapy with a β -lactam antibiotic may lower the risk of emergence of fluoroquinolone resistance, especially if *P. aeruginosa* infection is treated with ciprofloxacin or staphylococcal infection with the combination of a fluoroquinolone plus rifampin. These suggestions are based on observational studies and expert opinion. A recent randomized, controlled trial showed that 6 weeks of antibiotic treatment is not inferior to a 12-week course in patients with pyogenic vertebral osteomyelitis. The cure rate was 90.9% in both groups 1 year after

therapy. Thus, prolonged antibiotic therapy is required only for patients with undrained abscesses and for patients with spinal implants. Treatment efficacy should be regularly monitored through inquiries about signs and symptoms (fever, pain) and assessment for signs of inflammation (elevated CRP concentrations). Follow-up MRI is appropriate only for patients with pyogenic complications, since the correlation between clinical healing and improvement on MRI is very poor.

Surgical treatment generally is not needed in acute hematogenous vertebral osteomyelitis. However, it is always necessary in implant-associated spinal infection. Early infections (those occurring up to 30 days after internal stabilization) can be cured with debridement, implant retention, and a 3-month course of antibiotics (**Table 126-2**). In contrast, in late infection with a duration of >30 days, implant removal and a 6-week course of antibiotics (**Table 126-1**) are required for complete elimination of the infection. If implants cannot be removed, oral suppressive long-term treatment should follow the initial course of IV antibiotics. The optimal duration of suppressive therapy is unknown. However, if antibiotic therapy is discontinued after, for example, 1 year, close clinical and laboratory (CRP) follow-up is needed.

TABLE 126-1

Antibiotic Therapy for Osteomyelitis in Adults Without Implants^a

MICROORGANISM	ANTIMICROBIAL AGENT (DOSE, ^b ROUTE)
<i>Staphylococcus</i> spp.	
Methicillin-susceptible	Nafcillin or oxacillin ^c (2 g IV q6h)
	<i>followed by</i>
	Rifampin (300–450 mg PO q12h) plus levofloxacin (750 mg PO q24h or 500 mg PO q12h)
Methicillin-resistant	Vancomycin ^d (15 mg/kg IV q12h) or daptomycin (>6–8 mg/kg IV q24h)
	<i>followed by</i>
	Rifampin (300–450 mg PO q12h)
	<i>plus</i>
	Levofloxacin (750 mg PO q24h or 500 mg PO q12h) or TMP-SMX ^e (1 double-strength tablet PO q8h) or fusidic acid (500 mg PO q8h)
<i>Streptococcus</i> spp.	Penicillin G ^c (5 million units IV q6h) or ceftriaxone (2 g IV q24h)
Enterobacteriaceae	
Quinolone-susceptible	Ciprofloxacin (750 mg PO q24h)
Quinolone-resistant ^f	Imipenem (500 mg IV q6h)
<i>Pseudomonas aeruginosa</i>	Cefepime or ceftazidime (2 g IV q8h) <i>plus</i> an aminoglycoside ^g

MICROORGANISM	ANTIMICROBIAL AGENT (DOSE, ^b ROUTE)
	<i>or</i>
	Piperacillin-tazobactam (4.5 g IV q8h) plus an aminoglycoside ^g for 2–4 weeks
	<i>followed by</i>
	Ciprofloxacin ^h (750 mg PO q12h)
Anaerobes	Clindamycin (600 mg IV q6–8h) for 2–4 weeks
	<i>followed by</i>
	Clindamycin ⁱ (300 mg PO q6h)

a

Unless otherwise indicated, the total duration of antimicrobial treatment is generally 6 weeks.

b

All dosages are for adults with normal renal function.

c

When the patient has delayed-type penicillin hypersensitivity, cefuroxime (1.5 g IV q6–8h) can be administered. When the patient has immediate-type penicillin hypersensitivity, the penicillin should be replaced by vancomycin (1 g IV q12h).

d

Target vancomycin trough level: 15–20 µg/mL.

e

Trimethoprim-sulfamethoxazole. A double-strength tablet contains 160 mg of trimethoprim and 800 mg of sulfamethoxazole.

f

Including isolates producing extended-spectrum β-lactamase.

g

The need for addition of an aminoglycoside has not yet been proven. However, this addition may decrease the risk of emergence of resistance to the β-lactam.

h

The rationale for starting ciprofloxacin treatment only after pretreatment with a β-lactam is the increased risk of emergence of quinolone resistance in the presence of a heavy bacterial load.

i

Alternatively, penicillin G (5 million units IV q6h) or ceftriaxone (2 g IV q24h) can be used against gram-positive anaerobes (e.g., *Propionibacterium acnes*), and metronidazole (500 mg IV/PO q8h) can be used against gram-negative anaerobes (e.g., *Bacteroides* spp.).

Source: From W Zimmerli: N Engl J Med 362:1022, 2010. © Massachusetts Medical Society. Reprinted with permission.

TABLE 126-2

Antibiotic Therapy for Osteomyelitis Associated with Orthopedic Devices

MICROORGANISM	ANTIMICROBIAL AGENT ^a (DOSE, ROUTE)
<i>Staphylococcus</i> spp.	<i>Recommendation for initial treatment phase (2 weeks with implant)</i>
Methicillin-susceptible	Rifampin (450 mg PO/IV q12h ^b)
	<i>plus</i>
	Nafcillin or oxacillin ^c (2 g IV q6h)
Methicillin-resistant	Rifampin (450 mg PO/IV q12h ^b)
	<i>plus</i>
	Vancomycin (15 mg/kg IV q12h) or daptomycin (6–10 mg/kg IV q24h)
<i>Staphylococcus</i> spp.	<i>Recommendation after completion of initial treatment phase</i>
	Rifampin (450 mg PO q12h ^b)
	<i>plus</i>
	Levofloxacin (750 mg PO q24h or 500 mg PO q12h) or ciprofloxacin (750 mg PO q12h) or fusidic acid (500 mg PO q8h) or TMP-SMX ^d (1 double-strength tablet PO q8h) or minocycline (100 mg PO q12h) or linezolid (600 mg PO q12h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)
<i>Streptococcus</i> spp. ^e	Penicillin G ^c (18–24 million units/d IV in 6 divided doses) or ceftriaxone (2 g IV q24h) for 4 weeks
	<i>followed by</i>

MICROORGANISM	ANTIMICROBIAL AGENT ^a (DOSE, ROUTE)
	Amoxicillin (750–1000 mg PO q6–8h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)
<i>Enterococcus</i> spp. ^f	
Penicillin-susceptible	Penicillin G ^c (24 million units/d IV in 6 divided doses) <i>or</i> ampicillin or amoxicillin ^g (2 g IV q4–6h)
Penicillin-resistant	Vancomycin (15 mg/kg IV q12h) or daptomycin (6–10 mg/kg IV q24h) or linezolid (600 mg IV/PO q12h)
Enterobacteriaceae	A β -lactam selected in light of in vitro susceptibility profile for 2 weeks ^h
	<i>followed by</i>
	Ciprofloxacin (750 mg PO q12h)
<i>Enterobacter</i> spp. ⁱ and nonfermenters ^j (e.g., <i>Pseudomonas aeruginosa</i>)	Cefepime or ceftazidime (2 g IV q8h) <i>or</i> meropenem (1 g IV q8h ^k) for 2–4 weeks
	<i>followed by</i>
	Ciprofloxacin (750 mg PO q12h)
<i>Propionibacterium</i> spp.	Penicillin G ^c (18–24 million units/d IV in 6 divided doses) or clindamycin (600–900 mg IV q8h) for 2–4 weeks
	<i>followed by</i>
	Amoxicillin (750–1000 mg PO q6–8h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)

MICROORGANISM	ANTIMICROBIAL AGENT ^a (DOSE, ROUTE)
Gram-negative anaerobes (e.g., <i>Bacteroides</i> spp.)	Metronidazole (500 mg IV/PO q8h)
Mixed bacteria (without methicillin-resistant staphylococci)	Ampicillin-sulbactam (3 g IV q6h) or amoxicillin-clavulanate ^l (2.2 g IV q6h) or piperacillin-tazobactam (4.5 g IV q8h) or imipenem (500 mg IV q6h) or meropenem (1 g IV q8h ^k) for 2–4 weeks
	<i>followed by</i>
	Individualized oral regimens chosen in light of antimicrobial susceptibility

a

Antimicrobial agents should be chosen in light of the isolate's in vitro susceptibility, the patient's drug allergies and intolerances, potential drug interactions, and contraindications to specific drugs. All dosages recommended are for adults with normal renal and hepatic function. See text for total durations of antibiotic treatment.

b

Other dosages and intervals of administration with equivalent success rates have been reported.

c

When the patient has delayed-type penicillin hypersensitivity, cefazolin (2 g IV q8h) can be administered. When the patient has immediate-type penicillin hypersensitivity, the penicillin should be replaced by vancomycin (1 g IV q12h).

d

Trimethoprim-sulfamethoxazole. A double-strength tablet contains 160 mg of trimethoprim and 800 mg of sulfamethoxazole.

e

Determination of the minimal inhibitory concentration (MIC) of penicillin is advisable.

f

Combination therapy with an aminoglycoside is optional since its superiority to monotherapy for prosthetic joint infection is unproved. When using combination therapy, monitor for signs of aminoglycoside ototoxicity and nephrotoxicity; the latter is potentiated by other nephrotoxic agents (e.g., vancomycin).

g

For patients with hypersensitivity to penicillin, see treatment options for penicillin-resistant enterococci.

h

[Ciprofloxacin](#) (PO or IV) can be administered to patients with hypersensitivity to β -lactams.

i

Ceftriaxone and ceftazidime should not be administered for treatment targeting *Enterobacter* species, even strains that test susceptible in the laboratory, but can be used against nonfermenters. Strains producing extended-spectrum β -lactamases should not be treated with any cephalosporin, including cefepime. *Enterobacter* infections can also be treated with ertapenem (1 g IV q24h); however, ertapenem is not effective against *Pseudomonas* spp. and other nonfermenters.

j

Addition of an aminoglycoside is optional. Use of two active drugs can be considered in light of the patient's clinical condition.

k

The recommended dosage is in line with the guidelines of the Infectious Diseases Society of America. In Europe, 2 g IV q8h is suggested for *P. aeruginosa* infections.

l

Not available as an IV formulation in the United States.

Source: Modified from W Zimmerli et al: N Engl J Med 351:1645, 2004. © Massachusetts Medical Society. Reprinted with permission.

COMPLICATIONS

Complications should be suspected when there is persistent pain, a persistently increased CRP level, and new-onset or persistent neurologic impairment. In cases of persistent pain with or without signs of inflammation, paravertebral, epidural, or psoas abscesses (**Fig. 126-1**) must be sought. Epidural abscesses occur in 15–20% of cases. This complication is more common in the cervical column (30%) than in the lumbar spine (12%). Persistent pain despite normalization of CRP values indicates mechanical complications such as severe osteonecrosis or spinal instability. These patients require a consult with an experienced orthopedic surgeon.


FIGURE 126-1

CT scan of acute vertebral osteomyelitis (L1/L2) due to *Staphylococcus aureus* in a 64-year-old man. Low-grade fever persisted despite appropriate IV antibiotic therapy. The scan revealed a psoas abscess on the right side.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

GLOBAL CONSIDERATIONS

 The incidence rate of acute vertebral osteomyelitis is similar in different regions of the world. In contrast, subacute/chronic vertebral osteomyelitis predominates in defined regions. Cases attributable to brucellosis predominate in endemic areas such as the Middle East, Africa, Central and South America, and the Indian subcontinent. Tuberculosis is an especially frequent cause in Africa and Asia (India, Indonesia, China), where more than two-thirds of the global tuberculosis burden is reported. Thus, specific diagnostic tests are needed in patients either living in or having traveled to these regions.

OSTEOMYELITIS IN LONG BONES

PATHOGENESIS

Osteomyelitis in long bones is a consequence of hematogenous seeding, exogenous contamination during trauma (open fracture), or perioperative contamination during orthopedic repairs. Its presentation is either acute (with a duration of days to a few weeks) or chronic. Hematogenous infection in long bones typically occurs in children. Ineffectively treated hematogenous osteomyelitis during childhood can progress to chronic disease. In adults, the leading pathogenic source is exogenous infection, mainly associated with

internal fixation devices. Chronic osteomyelitis can recur after a symptom-free interval of >70 years. Such recurrences are most common among elderly patients who developed osteomyelitis in the preantibiotic era.

EPIDEMIOLOGY

In adults, most cases of long-bone osteomyelitis are posttraumatic or postsurgical; less frequently, late recurrence arises from hematogenous infections during childhood. The risk of infection depends on the type of fracture. After closed fracture, implant-associated infection occurs in fewer than 1% of patients. In contrast, after open fracture, the risk of osteomyelitis ranges from ~2% up to 30%, with the precise figure depending on the degree of tissue damage during trauma and the time between injury and admission to a specialized center.

MICROBIOLOGY

The spectrum of microorganisms causing hematogenous long-bone osteomyelitis does not differ from that in vertebral osteomyelitis. *S. aureus* is most commonly isolated in each type of osteomyelitis. In rare cases, mycobacteria or fungal agents such as *Cryptococcus* species, *Sporothrix schenckii*, *Blastomyces dermatitidis*, or *Coccidioides* species are found in patients who live or have traveled in endemic regions. Impaired cellular immunity (e.g., in HIV infection or after transplantation) predisposes to these etiologies. Coagulase-negative staphylococci are the second most common etiologic agents (after *S. aureus*) in implant-associated osteomyelitis. After open fracture, contiguous long-bone osteomyelitis is typically caused by gram-negative bacilli or a polymicrobial mixture of organisms.

CLINICAL MANIFESTATIONS

The leading symptoms in adults with primary or recurrent hematogenous long-bone osteomyelitis are pain and low-grade fever. Infection occasionally manifests as clinical sepsis and local signs of inflammation (erythema and swelling). After internal fixation, osteomyelitis can be classified as early (acute; <3 weeks), delayed (3–10 weeks), or late (chronic) infection. Early/acute long-bone osteomyelitis manifests as signs of surgical site infection, such as erythema and impaired wound healing. Acute implant-associated infection may also follow hematogenous seeding at any time after implantation of a device. Typical symptoms are new-onset pain and signs of sepsis. Delayed or late (chronic) infections are usually caused by low-virulence microorganisms or occur after ineffective treatment of early-onset infection. Patients may present with persisting pain, subtle local signs of inflammation, intermittent discharge of pus, or fluctuating erythema over the scar (**Fig. 126-2**).

FIGURE 126-2

A 42-year-old man who had had a malleolar fracture 6 weeks previously had persistent pain and slight inflammation after orthopedic repair. His infection was treated with oral antibiotics without debridement surgery. This insufficient management of an implant-associated *Staphylococcus aureus* infection was complicated by a sinus tract.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

DIAGNOSIS

The diagnostic workup for acute hematogenous long-bone osteomyelitis is similar to that for vertebral osteomyelitis. Bone remodeling and thus marker uptake are increased for at least 1 year after surgery. Therefore, the three-phase bone scan is not useful during this interval. However, in late recurrences it allows rapid diagnosis at low cost. If the results are positive, CT is required in order to estimate the extent of inflamed tissue and to detect bone necrosis (sequestrs). Implant-associated infection should be suspected if CRP values do not return to the normal range or rise after an initial decrease. Clinical and laboratory suspicion should prompt surgical exploration and sampling.

In chronic osteomyelitis of >1 year's duration, single-photon emission CT plus conventional CT (SPECT/CT) is a good option, either with ^{99m}Tc methylene diphosphonate (^{99m}Tc -MDP)-labeled leukocytes or with labeled monoclonal antibodies to granulocytes. Surgical debridement is needed for diagnostic (biopsy culture, histology) and therapeutic reasons.

TREATMENT

TREATMENT

Osteomyelitis in Long Bones

Treatment for acute hematogenous infection in long bones is identical to that for acute vertebral osteomyelitis (Table 126-1). The suggested duration of antibiotic therapy is 4–6 weeks. In contrast to chronic or implant-associated osteomyelitis, acute hematogenous infection does not generally require surgical

intervention. Initial IV administration of antimicrobial agents is followed by long-term oral treatment. The duration of the initial IV phase of therapy has not been defined. The IV course can be as short as 2 days for a drug with excellent bioavailability. In recurrences of chronic osteomyelitis as well as in each type of exogenous osteomyelitis (acute, chronic, with or without an implant), a combination of surgical debridement, obliteration of dead space, and long-term antibiotic therapy is needed.

The therapeutic aims in patients whose infections are associated with internal fixation devices are consolidation of the fracture and prevention of chronic osteomyelitis. Stable implants can be maintained except in patients with uncontrolled sepsis. Appropriate antimicrobial therapies are listed in [Table 126-2](#). The cure rate for early staphylococcal implant-associated infections treated with a fluoroquinolone plus rifampin is >90%. Rifampin is efficacious against staphylococcal biofilms of ≤ 3 weeks' duration. Similarly, fluoroquinolones are active against biofilms formed by gram-negative bacilli. In these cases, an initial 2-week course of IV therapy with a β -lactam is suggested in order to minimize the risk of emergence of resistance to the oral drugs. The total duration of treatment is 3 months, and the device can be retained even after antibiotics have been discontinued. In contrast, in cases caused by rifampin-resistant staphylococci or fluoroquinolone-resistant gram-negative bacilli, the hardware should be removed after consolidation of the fracture and before discontinuation of antibiotics. These patients are treated with an oral antibiotic (suppressive therapy) as long as retention of the hardware is necessary.

COMPLICATIONS

The main complication of long-bone osteomyelitis is the persistence of infection with progression to chronic osteomyelitis. This risk is especially high after internal fixation of an open fracture and among patients with implant-associated osteomyelitis that is treated without surgical debridement. In chronic osteomyelitis, recurrent sinus tracts result in severe damage to skin and soft tissue ([Fig. 126-2](#)). Patients who have chronic open wounds need a therapeutic approach combining orthopedic repair and plastic reconstructive surgery.

GLOBAL CONSIDERATIONS



In North American and Western European countries, tuberculous osteomyelitis is extremely rare, occurring mainly in very old people, in HIV-infected patients, and in immigrants from endemic countries. In contrast, in countries where the prevalence of tuberculosis is high (India, Indonesia, China), tuberculous osteomyelitis must routinely be considered.

PERIPROSTHETIC JOINT INFECTION

PATHOGENESIS

Implanted foreign material is highly susceptible to local infection due to local immunodeficiency around the device. Infection occurs by either the exogenous or the hematogenous route. More rarely, contiguous spread from adjacent sites of osteomyelitis or deep soft-tissue infection may cause periprosthetic joint infection

(PJI). The fact that foreign devices are covered with host proteins such as fibronectin favors the adherence of staphylococci and the formation of a biofilm that resists phagocytosis.

EPIDEMIOLOGY

The risk of infection manifesting during the first 2 postoperative years varies according to the joint. It is lowest after hip and knee arthroplasty (0.3–1.5%) and highest after ankle and elbow replacement (4–10%). The risk of hematogenous PJI is highest in the early postoperative period. However, hematogenous seeding occurs throughout life, and most cases therefore develop >2 years after implantation. The rate of risk for secondary PJI during *S. aureus* bacteremia is 30–40%.

MICROBIOLOGY

About 50–70% of cases of PJI are caused by staphylococci (*S. aureus* and coagulase-negative staphylococci), 6–10% by streptococci, 4–10% by gram-negative bacilli, and the rest by various other microorganisms. In some centers, the fraction of PJI cases caused by gram-negative bacilli is much higher for unknown reasons. All microorganisms can cause PJI, including fungi and mycobacteria. *P. acnes* causes up to one-third of episodes of periprosthetic shoulder infection.

CLASSIFICATION AND CLINICAL MANIFESTATIONS

PJI is traditionally classified as early (<3 months after implantation), delayed (3–24 months after surgery), or late (>2 years after implantation). For therapeutic decision-making (see below), it is more useful to classify PJI as (1) acute hematogenous PJI with <3 weeks of symptoms, (2) early postinterventional PJI manifesting within 1 month after surgery, or (3) chronic PJI with symptom duration of >3 weeks.

Acute exogenous PJI typically presents with local signs of infection (**Fig. 126-3**). In contrast, acute hematogenous PJI, most often caused by *S. aureus*, is characterized by new-onset pain that initially is not accompanied by prominent local inflammatory signs. In most cases, an ongoing sepsis syndrome dominates the clinical picture. Key findings in chronic PJI are joint effusion, local pain, implant loosening, and occasionally a sinus tract. Chronic PJI is most commonly caused by low-virulence microorganisms such as coagulase-negative staphylococci or *P. acnes*. These infections are characterized by nonspecific symptoms, such as chronic pain caused by low-grade inflammation or early loosening.

FIGURE 126-3

Acute postoperative periprosthetic joint infection of the left hip caused by group B streptococci in a 68-year-old woman.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

DIAGNOSIS

Blood tests such as the measurement of CRP (elevated levels, ≥ 10 mg/L) and erythrocyte sedimentation rate (elevated rates, ≥ 30 mm/h) are sensitive (91–97%) but not specific (70–78%). Synovial fluid cell counts are ~90% sensitive and specific, with threshold values of 1700 leukocytes/ μ L in periprosthetic knee infection and 4200 leukocytes/ μ L in periprosthetic hip infection. In the future, α -defensin, a biomarker that can be tested in synovial fluid, may replace cell counts. This test is expensive but accurate and easy to perform. During debridement surgery, at least three but optimally six tissue samples should be obtained for culture and histopathology. If implant material (modular parts, screws, or the prosthesis) is removed, sonication of this material followed by culture and/or use of molecular methods to examine the sonicate fluid allows the detection of microorganisms in biofilms.

The three-phase bone scan is very sensitive for detecting PJI but is not specific. As mentioned above, this test does not differentiate bone remodeling from infection and therefore is not useful during at least the first year after implantation. CT and MRI detect soft tissue infection, prosthetic loosening, and bone erosion, but imaging artifacts caused by metal implants limit their use. ^{18}F -fluorodeoxyglucose PET (^{18}F -FD-PET) is an alternative method with fair sensitivity and specificity for the detection of PJI. However, ^{18}F -FDG-PET/CT still is not an established technique for the diagnosis of PJI because of controversial published results.

TREATMENT

TREATMENT

Periprosthetic Joint Infection

Treatment of PJI requires a multidisciplinary approach involving an experienced orthopedic surgeon, an infectious disease specialist, a plastic reconstructive surgeon, and a microbiologist. Therefore, most patients are referred to a specialized center. In general, the goal of treatment is cure—i.e., a pain-free functional joint with complete eradication of the infecting pathogen(s). However, for patients with severe comorbidity, lifelong suppressive antimicrobial therapy may be preferred. As a rule, antimicrobial therapy without surgical intervention is not curative but merely suppressive. There are four curative surgical options: debridement and implant retention, one-stage implant exchange, two-stage implant exchange, and implant removal without replacement. Implant retention offers a good chance of infection-free survival (>80%) only if the following conditions are fulfilled: (1) acute infection, (2) stable implant, (3) pathogen susceptible to a biofilm-active antimicrobial agent (see below), and (4) skin and soft tissue in good condition.

Table 126-2 summarizes pathogen-specific antimicrobial therapy for PJI. Initial IV therapy is followed by long-term oral antibiotics. Efficacious treatment is best defined in staphylococcal implant-associated infections. Rifampin exhibits excellent activity against biofilms composed of susceptible staphylococci. Because of the risk of rapid emergence of resistance, rifampin must always be combined with another effective antibiotic. If gram-negative infections are treated with implant retention, fluoroquinolones should be used because of their activity against gram-negative biofilms.

PREVENTION OF HEMATOGENOUS INFECTION

As mentioned above, hematogenous seeding may occur throughout life. This risk is highest during *S. aureus* bacteremia from a distant focus. Therefore, documented bacterial infections should be promptly treated in patients with prosthetic joints. However, according to a large prospective case-control study, the risk of prosthetic hip or knee infection is not increased following dental procedures. Therefore, antibiotic prophylaxis is not needed during dental work.

GLOBAL CONSIDERATIONS



Rifampin and fluoroquinolones are still the only antimicrobial agents with good activity against staphylococcal and gram-negative biofilms, respectively. Thus, in countries with high rates of rifampin resistance in staphylococci and/or high rates of fluoroquinolone resistance in gram-negative bacilli, debridement with implant retention generally does not yield a good cure rate.

STERNAL OSTEOMYELITIS

PATHOGENESIS

Sternal osteomyelitis occurs primarily after sternal surgery (with the entry of exogenous organisms) and more rarely by hematogenous seeding or contiguous extension from adjacent sites of sternocostal arthritis. Exogenous sternal osteomyelitis after open sternal surgery is also called *deep sternal-wound infection*. Exogenous infection may follow minor sternal trauma, sternal fracture, and manubriosternal septic arthritis.

Tuberculous sternal osteomyelitis typically manifests during hematogenous seeding in children or as reactivated infection in adults. Reactivation is sometimes preceded by blunt trauma. In rare cases, tuberculous sternal osteomyelitis is caused by continuous infection from an infected internal mammary lymph node.

EPIDEMIOLOGY

The incidence of poststernotomy wound infection varies from 0.5% to 2%, but figures are even higher among patients with risk factors such as diabetes, obesity, chronic renal failure, emergency surgery, use of bilateral internal mammary arteries, and re-exploration for bleeding. Rapid diagnosis and correct management of superficial sternal wound infection prevent its progression to sternal osteomyelitis. Primary (hematogenous) sternal osteomyelitis accounts for only 0.3% of all cases of osteomyelitis. Risk factors are IV drug use, HIV infection, radiotherapy, blunt trauma, cardiopulmonary resuscitation, alcohol abuse, liver cirrhosis, and hemoglobinopathy.

MICROBIOLOGY

Poststernotomy osteomyelitis is generally caused by *S. aureus* (10–20% of cases), coagulase-negative staphylococci (40–60%), gram-negative bacilli (15–25%), or *P. acnes* (2–10%). Fungal infections caused by *Candida* species also play a role. The fact that ~20% of cases are polymicrobial is indicative of exogenous superinfection during therapy. Hematogenous sternal osteomyelitis is caused most commonly by *S. aureus*. Other microorganisms play a role in special populations—e.g., *P. aeruginosa* in IV drug users, *Salmonella* species in individuals with sickle cell anemia, and *M. tuberculosis* in patients from endemic areas who have previously had tuberculosis.

CLINICAL MANIFESTATIONS

Exogenous sternal osteomyelitis manifests as fever, increased local pain, erythema, wound discharge, and sternal instability (**Fig. 126-4**). Contiguous mediastinitis is a feared complication, occurring in ~10–30% of patients with sternal osteomyelitis. Hematogenous sternal osteomyelitis is characterized by sternal pain, swelling, and erythema. In addition, most patients have systemic signs and symptoms of sepsis.

FIGURE 126-4

Sternal osteomyelitis caused by *Staphylococcus epidermidis* 5 weeks after sternotomy for aortocoronary bypass in a 72-year-old man.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The differential diagnosis of hematogenous sternal osteomyelitis includes immunologic processes typically presenting as systemic or multifocal inflammation of the sternum or of the sternoclavicular or sternocostal joints (e.g., SAPHO [synovitis, acne, pustulosis, hyperostosis, osteitis], vasculitis, and chronic multifocal relapsing osteomyelitis).

DIAGNOSIS

In primary sternal osteomyelitis, the diagnostic workup does not differ from that in other types of hematogenous osteomyelitis (see above). When a patient has grown up in regions where tuberculosis is endemic, a specific workup for mycobacterial infection should be performed, especially if osteomyelitis had its onset after a blunt sternal trauma. In secondary sternal osteomyelitis, leukocyte counts may be normal, but the CRP level is >100 mg/L in most cases. Tissue sampling for microbiologic studies is crucial. In osteomyelitis associated with sternal wires, low-virulence microorganisms, such as coagulase-negative staphylococci, play an important role. In order to differentiate between colonization and infection, samples from at least three deep biopsies should be subjected to microbiologic examination. Superficial swab cultures are not diagnostic and may be misleading. No studies have compared the value of the various imaging modalities in suspected primary sternal osteomyelitis. However, MRI is the current gold standard for detection of each type of osteomyelitis.

TREATMENT

TREATMENT

Sternal Osteomyelitis

In cases of deep sternal-wound infection, antibiotic therapy should be started immediately after samples have been obtained for microbiologic analyses in order to control clinical sepsis. To protect a newly inserted


heart valve, initial treatment should be directed against staphylococci, with consideration of the local susceptibility pattern. In centers with a high prevalence of methicillin-resistant *S. aureus*, vancomycin or daptomycin should be added to a broad-spectrum β -lactam drug. As soon as cultures of blood and/or deep wound biopsies have confirmed the pathogen's identity and susceptibility pattern, treatment should be optimized and narrowed accordingly. [Tables 126-1](#) and [126-2](#) show appropriate therapeutic choices for the most frequently identified microorganisms causing sternal osteomyelitis in the absence and presence, respectively, of an implanted device. In a recent observational study of patients with staphylococcal deep sternal-wound infection, the use of a rifampin-containing regimen was predictive of success. The optimal duration of antibiotic therapy has not been established. In acute sternal osteomyelitis without hardware, a 6-week course is the rule. In patients with remaining sternal wires, treatment duration is generally prolonged to 3 months ([Table 126-2](#)). Like other types of tuberculous bone infection, tuberculous sternal osteomyelitis is treated for 6–12 months.

Primary sternal osteomyelitis can generally be treated without surgery. In contrast, in secondary sternal osteomyelitis, debridement is always required. This procedure should be performed by a team of experienced surgeons, since mediastinitis, bone infection, and skin and soft tissue damage may need to be treated during the same intervention.

PROGNOSIS

Primary sternal osteomyelitis poses a minimal mortality risk. In contrast, the in-hospital mortality rates from secondary sternal osteomyelitis are 15–30% after sternal surgery.

GLOBAL CONSIDERATIONS

 In endemic areas, microorganisms such as *M. tuberculosis*, *Salmonella* species, and *Brucella* species should be considered during sampling for microbiologic diagnosis.

FOOT OSTEOMYELITIS

PATHOGENESIS

Osteomyelitis of the foot usually occurs in patients with diabetes, peripheral arterial insufficiency, or peripheral neuropathy and after foot surgery. These entities are often linked to each other, especially in diabetic patients with late complications. However, foot osteomyelitis is also seen in patients with isolated peripheral neuropathy and can manifest as implant-associated osteomyelitis in patients without comorbidity due to a deep wound infection after foot surgery (hallux valgus surgery, arthrodesis, total ankle arthroplasty). Foot osteomyelitis is acquired almost exclusively by the exogenous route. It is a complication of deep pressure ulcers and of impaired wound healing after surgery.

EPIDEMIOLOGY

The incidence of diabetic foot infection is 30–40 cases/1000 persons with diabetes per year. The condition starts with skin and soft tissue lesions and progresses to osteomyelitis, especially in patients with risk factors. About 20–60% of patients with diabetic foot infection have confirmed osteomyelitis. Diabetic foot osteomyelitis increases the risk of amputation. With adequate management of the early stage of diabetic foot infections, the rate of amputation can be lowered.

RISK FACTORS

Risk factors for diabetic foot infection are (1) peripheral motor, sensory, and autonomic neuropathy; (2) neuro-osteoarthropathic deformities (Charcot foot; [Fig. 126-5](#)); (3) arterial insufficiency; (4) uncontrolled hyperglycemia; (5) disabilities such as reduced vision; and (6) maladaptive behavior.

FIGURE 126-5

Neuropathic joint disease (Charcot foot) complicated by chronic foot osteomyelitis in a 78-year old woman with diabetes mellitus complicated by severe neuropathy.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

MICROBIOLOGY

The correlation between cultures from bone biopsy and those from wound swabs or even deep soft-tissue punctures is poor. In a study of 31 patients with simultaneous sampling, the correlation between needle biopsy and bone biopsy cultures was only 24%. The correlation is better when *S. aureus* is isolated (40–50%) than when anaerobes (20–35%), gram-negative bacilli (20–30%), or coagulase-negative staphylococci (0–20%) are identified. When only bone-biopsy samples are considered, the leading pathogens are *S. aureus* (25–40%), anaerobes (5–20%), and various gram-negative bacilli (18–40%). The precise distribution depends

on whether the patient has already been treated with antibiotics. Anaerobes are especially prevalent in chronic wounds. Pretreatment typically selects for *P. aeruginosa*, methicillin-resistant *S. aureus*, or enterococci.

DIAGNOSIS

In many cases, foot osteomyelitis can be diagnosed clinically, without imaging procedures. Most clinicians rely on the “probe-to-bone” test, which has a positive predictive value of ~90% in populations with a high pretest probability. Thus, in a patient with diabetes who is hospitalized for a chronic deep foot ulcer, the diagnosis of foot osteomyelitis is highly probable if bone can be directly touched with a metal instrument. In a patient with a lower pretest probability, MRI should be performed because of its high degree of sensitivity (80–100%) and specificity (80–90%). Plain radiography has a sensitivity of only 30–90% and a specificity of only 50–90%; it may be considered for follow-up of patients with confirmed diabetic foot osteomyelitis.


TREATMENT

TREATMENT

Foot Osteomyelitis

As mentioned above, correlation between cultures of bone and those of wound swabs or wound punctures is poor. Antibiotic treatment should be based on bone culture. If no bone biopsy is performed, empirical therapy chosen in light of the most common infecting agents and the type of clinical syndrome should be given. Wound debridement combined with a 4- to 6-week course of antibiotics renders amputation unnecessary in about two-thirds of patients. According to the 2012 Infectious Diseases Society of America’s clinical practice guideline for the diagnosis and treatment of diabetic foot infections, the following management strategies should be considered. If a foot ulcer is clinically infected, prompt empirical antimicrobial therapy may prevent progression to osteomyelitis. When the risk of methicillin-resistant *S. aureus* is considered high, an agent active against these strains (e.g., vancomycin) should be chosen. If the patient has not recently received antibiotics, the spectrum of the selected antibiotic must include gram-positive cocci (e.g., [clindamycin](#), ampicillin-sulbactam). If the patient has received antibiotics within the past month, the spectrum of empirical antibiotics should include gram-negative bacilli (e.g., [clindamycin](#) plus a fluoroquinolone). If the patient has risk factors for *Pseudomonas* infection (previous colonization, residence in a warm climate, frequent exposure of the foot to water), an empirical antipseudomonal agent (e.g., piperacillin-tazobactam, cefepime) is indicated. If osteomyelitis is suspected either on clinical grounds (probe to bone) or on the basis of imaging procedures (MRI), bone biopsy should be performed. If infected bone is not entirely removed by surgery, the patient should be treated for 4–6 weeks in line with the identified pathogen(s) and their susceptibility. Treatment should initially be given by the IV route. Whether therapy can later be administered by the oral route depends on the bioavailability of oral drugs that cover the infecting agents. If dead bone cannot be removed, long-term therapy (at least 3 months) should be considered. In such cases, cure of osteomyelitis is usually the exception, and repetitive suppressive treatment may be needed.

GLOBAL CONSIDERATIONS

 The number of multiresistant microorganisms causing diabetic foot infection is increasing. The prevalence of methicillin-resistant *S. aureus* is 5–43% in various countries. In a study of 102 patients with diabetic foot infection from India, 69% of aerobic gram-negative bacilli produced extended-spectrum β -lactamase and 43% of *S. aureus* isolates were methicillin resistant. Risk factors for multidrug-resistant microorganisms are poor glycemic control, prolonged duration of infection, and large ulcer size.

FURTHER READING

Chan M et al: A retrospective study of deep sternal wound infections: Clinical and microbiological characteristics, treatment, and risk factors for complications. *Diag Microbiol Infect Dis* 84:261, 2016.

Lew DP, Waldvogel FA: Osteomyelitis. *Lancet* 364:369, 2004. [[PubMed: 15276398](#)]

Lipsky BA et al: 2012 Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. *Clin Infect Dis* 54:e132, 2012. [[PubMed: 22619242](#)]

Osmon DR et al: Diagnosis and management of prosthetic joint infection: Clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis* 56:e1, 2013. [[PubMed: 23223583](#)]

Zimmerli W: Vertebral osteomyelitis. *N Engl J Med* 362:1022, 2010. [[PubMed: 20237348](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 231: Approach to the Patient with Possible Cardiovascular Disease

Joseph Loscalzo

FIGURE 231-1

THE MAGNITUDE OF THE PROBLEM

Cardiovascular diseases comprise the most prevalent serious disorders in industrialized nations and are a rapidly growing problem in developing nations (**Chap. 233**). Age-adjusted death rates for coronary heart disease have declined by two-thirds in the last four decades in the United States, reflecting the identification and reduction of risk factors as well as improved treatments and interventions for the management of coronary artery disease, arrhythmias, and heart failure. Nonetheless, cardiovascular diseases remain the most common causes of death, responsible for 35% of all deaths, almost 1 million deaths each year. Approximately one-fourth of these deaths are sudden. In addition, cardiovascular diseases are highly prevalent, diagnosed in 80 million adults, or ~35% of the adult population. The growing prevalence of obesity (**Chap. 395**), type 2 diabetes mellitus (**Chap. 396**), and metabolic syndrome (**Chap. 401**), which are important risk factors for atherosclerosis, now threatens to reverse the progress that has been made in the age-adjusted reduction in the mortality rate of coronary heart disease.

For many years, cardiovascular disease was considered to be more common in men than in women. In fact, the percentage of all deaths secondary to cardiovascular disease is higher among women (43%) than among men (37%) (**Chap. 391**). In addition, although the absolute number of deaths secondary to cardiovascular disease has declined over the past decades in men, this number has actually risen in women. Inflammation, obesity, type 2 diabetes mellitus, and the metabolic syndrome appear to play more prominent roles in the development of coronary atherosclerosis in women than in men. Coronary artery disease (CAD) is more frequently associated with dysfunction of the coronary microcirculation in women than in men. Exercise electrocardiography has a lower diagnostic accuracy in the prediction of epicardial obstruction in women than in men.

NATURAL HISTORY

Cardiovascular disorders often present acutely, as in a previously asymptomatic person who develops an acute myocardial infarction (**Chap. 269**), or a previously asymptomatic patient with hypertrophic cardiomyopathy (**Chap. 254**) or with a prolonged QT interval (**Chap. 247**) whose first clinical manifestation is

syncope or even sudden death. However, the alert physician may recognize the patient at risk for these complications long before they occur and often can take measures to prevent their occurrence. For example, a patient with acute myocardial infarction will often have had risk factors for atherosclerosis for many years. Had these risk factors been recognized, their elimination or reduction might have delayed or even prevented the infarction. Similarly, a patient with hypertrophic cardiomyopathy may have had a heart murmur for years and a family history of this disorder. These findings could have led to an echocardiographic examination, recognition of the condition, and appropriate therapy long before the occurrence of a serious acute manifestation.

Patients with valvular heart disease or idiopathic dilated cardiomyopathy, by contrast, may have a prolonged course of gradually increasing dyspnea and other manifestations of chronic heart failure that is punctuated by episodes of acute deterioration only late in the course of the disease. Understanding the natural history of various cardiac disorders is essential for applying appropriate diagnostic and therapeutic measures to each stage of the condition, as well as for providing the patient and family with the likely prognosis.

CARDIAC SYMPTOMS

The symptoms caused by heart disease result most commonly from myocardial ischemia, disturbance of the contraction and/or relaxation of the myocardium, obstruction to blood flow, or an abnormal cardiac rhythm or rate. Ischemia, which is caused by an imbalance between the heart's oxygen supply and demand, is manifest most frequently as chest discomfort (**Chap. 11**), whereas reduction of the pumping ability of the heart commonly leads to fatigue and elevated intravascular pressure upstream of the failing ventricle. The latter results in abnormal fluid accumulation, with peripheral edema (**Chap. 37**) or pulmonary congestion and dyspnea (**Chap. 33**). Obstruction to blood flow, as occurs in valvular stenosis, can cause symptoms resembling those of myocardial failure (**Chap. 252**). Cardiac arrhythmias often develop suddenly, and the resulting symptoms and signs—palpitations (**Chap. 39**), dyspnea, hypotension, and syncope (**Chap. 18**)—generally occur abruptly and may disappear as rapidly as they develop.

Although dyspnea, chest discomfort, edema, and syncope are cardinal manifestations of cardiac disease, they occur in other conditions as well. Thus, dyspnea is observed in disorders as diverse as pulmonary disease, marked obesity, and anxiety (**Chap. 33**). Similarly, chest discomfort may result from a variety of noncardiac and cardiac causes other than myocardial ischemia (**Chap. 11**). Edema, an important finding in untreated or inadequately treated heart failure, also may occur with primary renal disease and in hepatic cirrhosis (**Chap. 37**). Syncope occurs not only with serious cardiac arrhythmias but in a number of neurologic conditions as well (**Chap. 18**). Whether heart disease is responsible for these symptoms frequently can be determined by carrying out a careful clinical examination (**Chap. 234**), supplemented by noninvasive testing using electrocardiography at rest and during exercise (**Chap. 235**), echocardiography, roentgenography, and other forms of myocardial imaging (**Chap. 236**).

Myocardial or coronary function that may be adequate at rest may be insufficient during exertion. Thus, dyspnea and/or chest discomfort that appear during activity are characteristic of patients with heart disease,

whereas the opposite pattern, that is, the appearance of these symptoms at rest and their remission during exertion, is rarely observed in such patients. It is important, therefore, to question the patient carefully about the relation of symptoms to exertion.

Many patients with cardiovascular disease may be asymptomatic both at rest and during exertion but may present with an abnormal physical finding such as a heart murmur, elevated arterial pressure, or an abnormality of the electrocardiogram (ECG) or imaging test. It is important to assess the global risk of CAD in asymptomatic individuals, using a combination of clinical assessment and measurement of cholesterol and its fractions, as well as other biomarkers, such as C-reactive protein, in some patients. Since the first clinical manifestation of CAD may be catastrophic—sudden cardiac death, acute myocardial infarction, or stroke in previous asymptomatic persons—it is mandatory to identify those at high risk of such events and institute further testing and preventive measures.

DIAGNOSIS

As outlined by the New York Heart Association (NYHA), the elements of a complete cardiac diagnosis include the systematic consideration of the following:

1. *The underlying etiology.* Is the disease congenital, hypertensive, ischemic, or inflammatory in origin?
2. *The anatomic abnormalities.* Which chambers are involved? Are they hypertrophied, dilated, or both? Which valves are affected? Are they regurgitant and/or stenotic? Is there pericardial involvement? Has there been a myocardial infarction?
3. *The physiologic disturbances.* Is an arrhythmia present? Is there evidence of congestive heart failure or myocardial ischemia?
4. *Functional disability.* How strenuous is the physical activity required to elicit symptoms? The classification provided by the NYHA has been found to be useful in describing functional disability ([Table 231-1](#)).

TABLE 231-1

New York Heart Association Functional Classification

Class I	Class III
No limitation of physical activity	Marked limitation of physical activity
No symptoms with ordinary exertion	Less than ordinary activity causes symptoms
Class II	Asymptomatic at rest
Slight limitation of physical activity	Class IV
Ordinary activity causes symptoms	Inability to carry out any physical activity without discomfort
	Symptoms at rest

Source: Data from The Criteria Committee of the New York Heart Association.

One example may serve to illustrate the importance of establishing a complete diagnosis. In a patient who presents with exertional chest discomfort, the identification of myocardial ischemia as the etiology is of great clinical importance. However, the simple recognition of ischemia is insufficient to formulate a therapeutic strategy or prognosis until the underlying anatomic abnormalities responsible for the myocardial ischemia, for example, coronary atherosclerosis or aortic stenosis, are identified and a judgment is made about whether other physiologic disturbances that cause an imbalance between myocardial oxygen supply and demand, such as severe anemia, thyrotoxicosis, or supraventricular tachycardia, play contributory roles. Finally, the severity of the disability should govern the extent and tempo of the workup and strongly influence the therapeutic strategy that is selected.

The establishment of a correct and complete cardiac diagnosis usually commences with the history and physical examination (**Chap. 234**). Indeed, the clinical examination remains the basis for the diagnosis of a wide variety of disorders. The clinical examination may then be supplemented by five types of laboratory tests: (1) ECG (**Chap. 235**), (2) noninvasive imaging examinations (chest roentgenogram, echocardiogram, radionuclide imaging, computed tomographic imaging, positron emission tomography, and magnetic resonance imaging) (**Chap. 236**), (3) blood tests to assess risk (e.g., lipid determinations, C-reactive protein) or cardiac function (e.g., brain natriuretic peptide [BNP] [**Chap. 252**]), (4) occasionally specialized invasive examinations (i.e., cardiac catheterization and coronary arteriography [**Chap. 237**]), and (5) genetic tests to identify monogenic cardiac diseases (e.g., hypertrophic cardiomyopathy [**Chap. 254**], Marfan's syndrome [**Chap. 406**], and abnormalities of cardiac ion channels that lead to prolongation of the QT interval and an increase in the risk of sudden death [**Chap. 241**]). These tests are becoming more widely available.

FAMILY HISTORY

In eliciting the history of a patient with known or suspected cardiovascular disease, particular attention should be directed to the family history. Familial clustering is common in many forms of heart disease. Mendelian transmission of single-gene defects may occur, as in hypertrophic cardiomyopathy ([Chap. 254](#)), Marfan's syndrome ([Chap. 406](#)), and sudden death associated with a prolonged QT syndrome ([Chap. 247](#)). Premature coronary disease and essential hypertension, type 2 diabetes mellitus, and hyperlipidemia (the most important risk factors for CAD) are usually polygenic disorders. Although familial transmission may be less obvious than in the monogenic disorders, it is helpful in assessing risk and prognosis in polygenic disorders, as well. Familial clustering of cardiovascular diseases not only may occur on a genetic basis but also may be related to familial dietary or behavior patterns, such as excessive ingestion of salt or calories and cigarette smoking.

ASSESSMENT OF FUNCTIONAL IMPAIRMENT

When an attempt is made to determine the severity of functional impairment in a patient with heart disease, it is helpful to ascertain the level of activity and the rate at which it is performed before symptoms develop. Thus, it is not sufficient to state that the patient complains of dyspnea. The breathlessness that occurs after running up two long flights of stairs denotes far less functional impairment than do similar symptoms that occur after taking a few steps on level ground. In addition, the degree of customary physical activity at work and during recreation should be considered. The development of two-flight dyspnea in a well-conditioned marathon runner may be far more significant than the development of one-flight dyspnea in a previously sedentary person. The history should include a detailed consideration of the patient's therapeutic regimen. For example, the persistence or development of edema, breathlessness, and other manifestations of heart failure in a patient who is receiving optimal doses of diuretics and other therapies for heart failure ([Chap. 252](#)) is far graver than are similar manifestations in the absence of treatment. Similarly, the presence of angina pectoris despite treatment with optimal doses of multiple antianginal drugs ([Chap. 267](#)) is more serious than it is in a patient on no therapy. In an effort to determine the progression of symptoms, and thus the severity of the underlying illness, it may be useful to ascertain what, if any, specific tasks the patient could have carried out 6 months or 1 year earlier that he or she cannot carry out at present.

ELECTROCARDIOGRAM

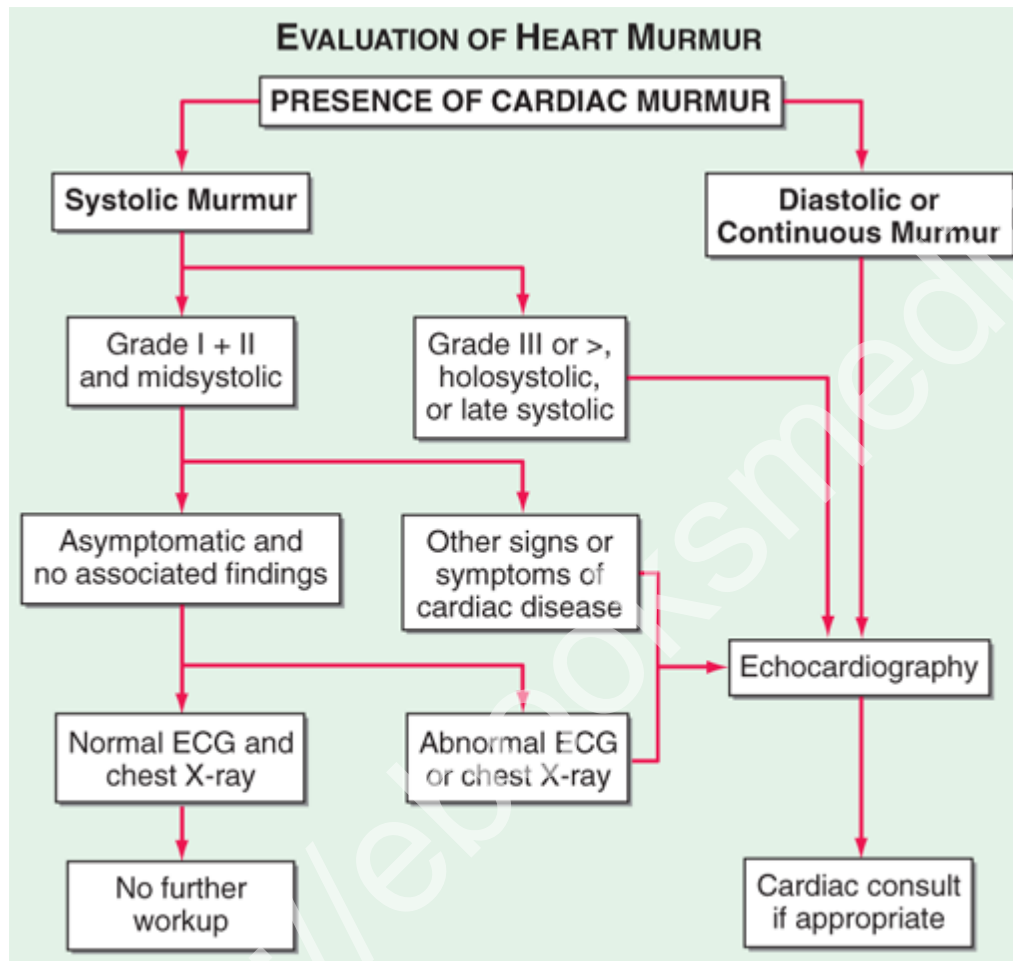
(See also [Chap. 235](#)) Although an ECG usually should be recorded in patients with known or suspected heart disease, with the exception of the identification of arrhythmias, conduction abnormalities, ventricular hypertrophy, and acute myocardial infarction, it generally does not establish a specific diagnosis. The range of normal electrocardiographic findings is wide, and the tracing can be affected significantly by many noncardiac factors, such as age, body habitus, and serum electrolyte concentrations. In general, electrocardiographic changes should be interpreted in the context of other abnormal cardiovascular findings.

ASSESSMENT OF THE PATIENT WITH A HEART MURMUR

(Fig. 231-1) The cause of a heart murmur can often be readily elucidated from a systematic evaluation of its major attributes: timing, duration, intensity, quality, frequency, configuration, location, and radiation when considered in the light of the history, general physical examination, and other features of the cardiac examination, as described in Chap. 234.

FIGURE 231-1

Approach to the evaluation of a heart murmur. ECG, electrocardiogram. (Reproduced with permission from *Primary Cardiology*, 2nd ed, E Braunwald, L Goldman [eds]. Philadelphia, Saunders, 2003.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: *Harrison's Principles of Internal Medicine*, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The majority of heart murmurs are midsystolic and soft (grades I–II/VI). When such a murmur occurs in an asymptomatic child or young adult *without* other evidence of heart disease on clinical examination, it is usually benign and echocardiography generally is not required. By contrast, two-dimensional and Doppler echocardiography (Chap. 236) are indicated in patients with loud systolic murmurs (grades \geq III/VI), especially those that are holosystolic or late systolic, and in most patients with diastolic or continuous murmurs.

PITFALLS IN CARDIOVASCULAR MEDICINE

Increasing subspecialization in internal medicine and the perfection of advanced diagnostic techniques in cardiology can lead to several undesirable consequences. Examples include the following:

1. Failure by the *noncardiologist* to recognize important cardiac manifestations of systemic illnesses. For example, the presence of mitral stenosis, patent foramen ovale, and/or transient atrial arrhythmia should be considered in a patient with stroke, or the presence of pulmonary hypertension and cor pulmonale should be considered in a patient with scleroderma or Raynaud's syndrome. A cardiovascular examination should be carried out to identify and estimate the severity of the cardiovascular involvement that accompanies many noncardiac disorders.
2. Failure by the *cardiologist* to recognize underlying systemic disorders in patients with heart disease. For example, hyperthyroidism should be considered in an elderly patient with atrial fibrillation and unexplained heart failure, and Lyme disease should be considered in a patient with unexplained fluctuating atrioventricular block. A cardiovascular abnormality may provide the clue critical to the recognition of some systemic disorders. For example, an unexplained pericardial effusion may provide an early clue to the diagnosis of tuberculosis or a neoplasm.
3. Overreliance on and overutilization of laboratory tests, particularly invasive techniques, for the evaluation of the cardiovascular system. Cardiac catheterization and coronary arteriography ([Chap. 237](#)) provide precise diagnostic information that may be crucial in developing a therapeutic plan in patients with known or suspected CAD. Although a great deal of attention has been directed to these examinations, it is important to recognize that they serve to *supplement*, not *supplant*, a careful examination carried out with clinical and noninvasive techniques. A coronary arteriogram should not be performed in lieu of a careful history in patients with chest pain suspected of having ischemic heart disease. Although coronary arteriography may establish whether the coronary arteries are obstructed and to what extent, the results of the procedure by themselves often do not provide a definitive answer to the question of whether a patient's complaint of chest discomfort is attributable to coronary atherosclerosis and whether or not revascularization is indicated.

Despite the value of invasive tests in certain circumstances, they entail some small risk to the patient, involve discomfort and substantial cost, and place a strain on medical facilities. Therefore, they should be carried out only if the results can be expected to modify the patient's management.

DISEASE PREVENTION AND MANAGEMENT

The prevention of heart disease, especially of CAD, is one of the most important tasks of primary health care givers as well as cardiologists. Prevention begins with risk assessment, followed by attention to lifestyle, such as achieving optimal weight, physical activity, and smoking cessation, and then aggressive treatment of all abnormal risk factors, such as hypertension, hyperlipidemia, and diabetes mellitus ([Chap. 396](#)).

After a complete diagnosis has been established in patients with known heart disease, a number of management options are usually available. Several examples may be used to demonstrate some of the

principles of cardiovascular therapeutics:

1. In the absence of evidence of heart disease, the patient should be clearly informed of this assessment and *not* be asked to return at intervals for repeated examinations. If there is no evidence of disease, such continued attention may lead to the patient's developing inappropriate concern about the possibility of heart disease.
2. If there is no evidence of cardiovascular disease but the patient has one or more risk factors for the development of ischemic heart disease (**Chap. 267**), a plan for their reduction should be developed and the patient should be retested at intervals to assess compliance and efficacy in risk reduction.
3. Asymptomatic or mildly symptomatic patients with valvular heart disease that is anatomically severe should be evaluated periodically, every 6 to 12 months, by clinical and noninvasive examinations. Early signs of deterioration of ventricular function may signify the need for surgical treatment before the development of disabling symptoms, irreversible myocardial damage, and excessive risk of surgical treatment (**Chap. 256**).
4. In patients with CAD (**Chap. 267**), available practice guidelines should be considered in the decision on the form of treatment (medical, percutaneous coronary intervention, or surgical revascularization). Mechanical revascularization may be employed too frequently in the United States and too infrequently in Eastern Europe and developing nations. The mere presence of angina pectoris and/or the demonstration of critical coronary arterial narrowing at angiography should not reflexively evoke a decision to treat the patient by revascularization. Instead, these interventions should be limited to patients with CAD whose angina has not responded adequately to medical treatment or in whom revascularization has been shown to improve the natural history (e.g., acute coronary syndrome or multivessel CAD with left ventricular dysfunction).

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

Silverchair

Chapter 278: Approach to the Patient with Disease of the Respiratory System

Patricia A. Kritek; Bruce D. Levy

INTRODUCTION

The majority of diseases of the respiratory system present with cough and/or dyspnea and fall into one of three major categories: (1) obstructive lung diseases; (2) restrictive disorders; and (3) abnormalities of the vasculature. Obstructive lung diseases are most common and primarily disorders of the airways, such as asthma, chronic obstructive pulmonary disease (COPD), bronchiectasis, and bronchiolitis. Diseases resulting in restrictive pathophysiology include parenchymal lung diseases, abnormalities of the chest wall and pleura, and neuromuscular disease. Pulmonary embolism, pulmonary hypertension, and pulmonary venoocclusive disease are all disorders of the pulmonary vasculature. Although many specific diseases fall into these major categories, both infective and neoplastic processes can affect the respiratory system and result in myriad pathologic findings, including those listed in the three categories above ([Table 278-1](#)).

TABLE 278-1

Categories of Respiratory Disease

Category	Examples
Obstructive lung disease	Asthma Chronic obstructive pulmonary disease (COPD) Bronchiectasis Bronchiolitis
Restrictive pathophysiology—parenchymal disease	Idiopathic pulmonary fibrosis (IPF) Asbestosis Desquamative interstitial pneumonitis (DIP) Sarcoidosis
Restrictive pathophysiology—neuromuscular weakness	Amyotrophic lateral sclerosis (ALS) Guillain-Barré syndrome Myasthenia gravis
Restrictive pathophysiology—chest wall/pleural disease	Kyphoscoliosis Ankylosing spondylitis Chronic pleural effusions
Pulmonary vascular disease	Pulmonary embolism Pulmonary arterial hypertension (PAH) Pulmonary venoocclusive disease Vasculitis
Malignancy	Bronchogenic carcinoma (non-small-cell and small-cell lung cancer) Metastatic disease
Infectious diseases	Pneumonia Bronchitis Tracheitis

Disorders can also be grouped according to gas exchange abnormalities, including hypoxemic, hypercarbic, or combined impairment; however, many respiratory disorders do not manifest as gas exchange abnormalities.

As with the evaluation of most patients, the approach to a patient with a respiratory system disorder begins with a thorough history and a focused physical examination. Many patients will subsequently undergo pulmonary function testing, chest imaging, blood and sputum analysis, a variety of serologic or microbiologic studies, and diagnostic procedures, such as bronchoscopy. This stepwise approach is discussed in detail below.

HISTORY

Dyspnea and Cough

The cardinal symptoms of respiratory disease are dyspnea and cough (**Chaps. 33 and 34**). Dyspnea has many causes, some of which are not predominantly due to lung pathology. The words a patient uses to describe shortness of breath can suggest certain etiologies for dyspnea. Patients with obstructive lung disease often complain of “chest tightness” or “inability to get a deep breath,” whereas patients with congestive heart failure more commonly report “air hunger” or a sense of suffocation.

The tempo of onset and the duration of a patient’s dyspnea are likewise helpful in determining the etiology. Acute shortness of breath is usually associated with sudden physiological changes, such as laryngeal edema, bronchospasm, myocardial infarction, pulmonary embolism, or pneumothorax. Patients with COPD and idiopathic pulmonary fibrosis (IPF) experience a gradual progression of dyspnea on exertion, punctuated by acute exacerbations of shortness of breath. In contrast, most asthmatics do not have daily symptoms, but experience intermittent episodes of dyspnea, cough, and chest tightness that are usually associated with specific triggers, such as an upper respiratory tract infection or exposure to allergens.

Specific questioning should focus on factors that incite dyspnea as well as on any intervention that helps resolve the patient’s shortness of breath. Asthma is commonly exacerbated by specific triggers, although this can also be true of COPD. Many patients with lung disease report dyspnea on exertion. Determining the degree of activity that results in shortness of breath gives the clinician a gauge of the patient’s degree of disability. Many patients adapt their level of activity to accommodate progressive limitation. For this reason, it is important, particularly in older patients, to delineate the activities in which they engage and how these activities have changed over time. Dyspnea on exertion is often an early symptom of underlying lung or heart disease and warrants a thorough evaluation.

For cough, the clinician should inquire about the duration of the cough, whether or not it is associated with sputum production, and any specific triggers that induce it. Acute cough productive of phlegm is often a symptom of infection of the respiratory system, including processes affecting the upper airway (e.g., sinusitis, tracheitis), the lower airways (e.g., bronchitis, bronchiectasis), and the lung parenchyma (e.g., pneumonia). Both the quantity and quality of the sputum, including whether it is blood-streaked or frankly bloody, should be determined. Hemoptysis warrants urgent evaluation as delineated in **Chap. 35**.

Chronic cough (defined as that persisting for >8 weeks) is commonly associated with obstructive lung diseases, particularly asthma, COPD and chronic bronchiectasis, as well as “nonrespiratory” diseases, such as gastroesophageal reflux and postnasal drip. Diffuse parenchymal lung diseases, including IPF, frequently

present as a persistent, nonproductive cough. All causes of cough are not respiratory in origin, and assessment should encompass a broad differential, including cardiac and gastrointestinal diseases as well as psychogenic causes.

Additional Symptoms

Patients with respiratory disease may report wheezing, which is suggestive of airways disease, particularly asthma. Hemoptysis can be a symptom of a variety of lung diseases, including infections of the respiratory tract, bronchogenic carcinoma, and pulmonary embolism. In addition, chest pain or discomfort can be respiratory in origin. As the lung parenchyma is not innervated with pain fibers, pain in the chest from respiratory disorders usually results from either diseases of the parietal pleura (e.g., pneumothorax) or pulmonary vascular diseases (e.g., pulmonary hypertension). As many diseases of the lung can result in strain on the right side of the heart, patients may also present with symptoms of cor pulmonale, including abdominal bloating or distention and pedal edema ([Chap. 252](#)).

Additional History

A thorough social history is an essential component of the evaluation of patients with respiratory disease. All patients should be asked about current or previous cigarette smoking, as this exposure is associated with many diseases of the respiratory system, including COPD, bronchogenic lung cancer, and select parenchymal lung diseases (e.g., desquamative interstitial pneumonitis and pulmonary Langerhans cell histiocytosis). For most of these disorders, increased cigarette smoke exposure (i.e., cigarette pack-years) increases the risk of disease. “Second-hand smoke” also increases risk for some respiratory disorders, so patients should also be asked about parents, spouses, or housemates who smoke. Possible inhalational exposures at work (e.g., asbestos, silica) or home (e.g., wood smoke, excrement from pet birds) should be explored ([Chap. 283](#)). Travel predisposes to certain infections of the respiratory tract, most notably tuberculosis. Potential exposure to fungi is increased in specific geographic regions or climates (e.g., *Histoplasma capsulatum*), so exposures to these regions should be determined.

Associated symptoms of fever and chills should raise the suspicion of infective etiologies, both pulmonary and systemic. A comprehensive review of systems may suggest rheumatologic or autoimmune disease presenting with respiratory tract manifestations. Questions should focus on joint pain or swelling, rashes, dry eyes, dry mouth, or constitutional symptoms. In addition, carcinomas from a variety of primary sources commonly metastasize to the lung and cause respiratory symptoms. Finally, therapy for other conditions, including both irradiation and medications, can result in diseases of the chest.

Physical Examination

The clinician’s suspicion of respiratory disease often begins with patient’s vital signs. The respiratory rate is informative, whether elevated (tachypnea) or depressed (hypopnea). In addition, pulse oximetry should be measured, as many patients with respiratory disease have hypoxemia, either at rest or with exertion.

The first step of the physical examination is inspection. Patients with respiratory disease may be in distress, using accessory muscles of respiration to breathe. Severe kyphoscoliosis can result in restrictive pathophysiology. Inability to complete a sentence in conversation is generally a sign of severe impairment and should result in an expedited evaluation of the patient.

Percussion of the chest is used to establish diaphragm excursion and lung size. In the setting of decreased breath sounds, percussion is used to distinguish between pleural effusions (dull to percussion) and pneumothorax (hyper-resonant note).

The role of palpation is limited in the respiratory examination. Palpation can demonstrate subcutaneous air in the setting of barotrauma. It can also be used as an adjunctive assessment to determine whether an area of decreased breath sounds is due to consolidation (increased tactile fremitus) or a pleural effusion (decreased tactile fremitus). To detect unilateral disorders of ventilation, the symmetry and degree of chest wall expansion can be assessed during a deep inspiration by placing one's thumbs together at the midline over the lower posterior chest while grasping the lateral rib cage.

The majority of the manifestations of respiratory disease present as abnormalities of auscultation. Wheezes are a manifestation of airway obstruction. While most commonly a sign of asthma, peribronchial edema in the setting of congestive heart failure can also result in diffuse wheezes, as can any other process that causes narrowing of small airways. Wheezes can be polyphonic, involving multiple different size airways (e.g., asthma) or monophonic, involving one size airway (e.g., bronchogenic carcinoma). For these reasons, clinicians must take care not to attribute all wheezing to asthma.

Rhonchi are a manifestation of obstruction of medium-sized airways, most often with secretions. In the acute setting, this manifestation may be a sign of viral or bacterial bronchitis. Chronic rhonchi suggest bronchiectasis or COPD. In contrast to expiratory wheezes and rhonchi, stridor is a high-pitched, focal inspiratory wheeze, usually heard over the neck as a manifestation of upper airway obstruction.

Crackles, or rales, are commonly a sign of alveolar disease. Processes that fill the alveoli with fluid may result in crackles, including pulmonary edema and pneumonia. Crackles in pulmonary edema are generally more prominent at the bases. Interestingly, diseases that result in fibrosis of the interstitium (e.g., IPF) also result in crackles that sound like Velcro being ripped apart. Although some clinicians make a distinction between "wet" and "dry" crackles, this distinction has not been shown to be a reliable way to differentiate among etiologies of respiratory disease.

One way to help distinguish between crackles associated with alveolar fluid and those associated with interstitial fibrosis is to assess for egophony. *Egophony* is the auscultation of the sound "AH" instead of "EEE" when a patient phonates "EEE." This change in note is due to abnormal sound transmission through consolidated parenchyma and is present in pneumonia but not in IPF. Similarly, areas of alveolar filling have increased whispered *pectoriloquy* as well as transmission of larger-airway sounds (i.e., bronchial breath sounds in a lung zone where vesicular breath sounds are expected).

The lack or diminution of breath sounds can also help determine the etiology of respiratory disease. Patients with emphysema often have a quiet chest with diffusely decreased breath sounds. A pneumothorax or pleural effusion may present with an area of absent breath sounds.

Pedal edema, if symmetric, may suggest cor pulmonale; if asymmetric, it may be due to deep venous thrombosis and associated pulmonary embolism. Jugular venous distention may also be a sign of volume overload associated with right heart failure. *Pulsus paradoxus* is an ominous sign in a patient with obstructive lung disease, as it is associated with significant negative intrathoracic (pleural) pressures required for ventilation and impending respiratory failure.

As stated earlier, rheumatologic disease may manifest primarily as lung disease. Owing to this association, particular attention should be paid to joint and skin examination. Clubbing can be found in many lung diseases, including cystic fibrosis, IPF, and lung cancer. Cyanosis is seen in hypoxemic respiratory disorders that result in >5 g of deoxygenated hemoglobin/dL.

DIAGNOSTIC EVALUATION

The sequence of studies is dictated by the clinician's differential diagnosis, as determined by the history and physical examination. Acute respiratory symptoms are often evaluated with multiple tests performed at the same time in order to diagnose any life-threatening diseases rapidly (e.g., pulmonary embolism or multilobar pneumonia). In contrast, chronic dyspnea and cough can be evaluated in a more protracted, stepwise fashion.

Pulmonary Function Testing

The initial pulmonary function test obtained is spirometry (See also [Chap. 280](#)). This study is an effort-dependent test used to assess for obstructive pathophysiology as seen in asthma, COPD, and bronchiectasis. A diminished-forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) (often defined as $<70\%$) is diagnostic of obstruction. In addition to measuring FEV_1 and FVC, the clinician should examine the flow-volume loop (which is less effort-dependent). A plateau of the inspiratory and expiratory curves suggests large-airway obstruction in extrathoracic and intrathoracic locations, respectively.

Spirometry with symmetric decreases in FEV_1 and FVC warrants further testing, including measurement of lung volumes and the diffusion capacity of the lung for carbon monoxide (D_LCO). A total lung capacity $<80\%$ of the patient's predicted value defines restrictive pathophysiology. Restriction can result from parenchymal disease, neuromuscular weakness, or chest wall or pleural diseases ([Table 278-1](#)). Restriction with impaired gas exchange, as indicated by a decreased D_LCO , suggests parenchymal lung disease. Additional testing, such as measurements of maximal inspiratory and expiratory pressures, can help diagnose neuromuscular weakness. Normal spirometry, normal lung volumes, and a low D_LCO should prompt further evaluation for pulmonary vascular disease.

Arterial blood gas testing is often helpful in assessing respiratory disease. Hypoxemia, while usually apparent with pulse oximetry, can be further evaluated with the measurement of arterial PO_2 and the calculation of an alveolar gas and arterial blood oxygen tension difference ($[A-a]DO_2$). Patients with diseases that cause ventilation-perfusion mismatch or shunt physiology have an increased $(A-a)DO_2$ at rest. Arterial blood gas

testing also allows the measurement of arterial PCO₂. Hypercarbia can accompany disorders of ventilation, as seen in severe airway obstruction (e.g., COPD) or progressive restrictive physiology.

Chest Imaging

Most patients with disease of the respiratory system undergo imaging of the chest as part of the initial evaluation (See [Chap. A12](#)). Clinicians should generally begin with ultrasound of the chest or a plain chest radiograph, preferably posterior-anterior and lateral films. Ultrasound is often readily available and can help rapidly diagnose pneumothorax, pleural effusion, and consolidation of lung parenchyma. Chest radiographs give additional detail and can reveal findings including opacities of the parenchyma, blunting of the costophrenic angles, mass lesions, and volume loss. However, many diseases of the respiratory system, particularly those of the airways and pulmonary vasculature, are associated with a normal chest radiograph.

CT scan of the chest can also be useful to delineate parenchymal processes, pleural disease, masses or nodules, and large airways. If the test includes administration of contrast, the pulmonary vasculature can be assessed with particular utility for determination of pulmonary emboli. Intravenous contrast also allows lymph nodes to be examined in greater detail. When coupled with positron emission testing (PET), lesions of the chest can be assessed for metabolic activity; helping differentiate between malignancy and scar.

FURTHER STUDIES

Depending on the clinician's suspicion, a variety of other studies may be done. Concern about large-airway lesions may warrant bronchoscopy. This procedure may also be used to sample the alveolar space with bronchoalveolar lavage or to obtain nonsurgical lung biopsies. Blood testing may include assessment for hypercoagulable states in the setting of pulmonary vascular disease, serologic testing for infectious or rheumatologic disease, or assessment of inflammatory markers or leukocyte counts (e.g., eosinophils). Genetic testing is increasingly used for heritable lung diseases such as cystic fibrosis. Sputum evaluation for malignant cells or microorganisms may be appropriate. An echocardiogram to assess right- and left-sided heart function is often obtained. Finally, at times, a surgical lung biopsy is needed to diagnose certain diseases of the respiratory system. All of these studies will be guided by the preceding history, physical examination, pulmonary function testing, and chest imaging.

FURTHER READING

Achilleos A: Evidence-based evaluation and management of chronic cough. *Med Clin North Am* 100:1033, 2016. [[PubMed: 27542423](#)]

Bohadana A et al: Fundamentals of lung auscultation. *N Engl J Med* 370:744, 2014. [[PubMed: 24552321](#)]

Koenig SJ et al: Thoracic ultrasonography for the pulmonary specialist. *Chest* 140:1332, 2011. [[PubMed: 22045878](#)]

Parshall MB et al: An official American Thoracic Society statement: Update on the mechanisms, assessment, and management of dyspnea. Am J Respir Crit Care Med 185:435, 2012. [[PubMed: 22336677](#)]

Pellegrino R et al: Interpretive strategies for lung function tests. Eur Respir J 26:948, 2005. [[PubMed: 16264058](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 279: Disturbances of Respiratory Function

Edward T. Naureckas; Julian Solway

INTRODUCTION

The primary functions of the respiratory system—to oxygenate blood and eliminate carbon dioxide—require virtual contact between blood and fresh air, which facilitates diffusion of respiratory gases between blood and gas. This process occurs in the lung alveoli, where blood flowing through alveolar wall capillaries is separated from alveolar gas by an extremely thin membrane of flattened endothelial and epithelial cells, across which respiratory gases diffuse and equilibrate. Blood flow through the lung is unidirectional via a continuous vascular path along which venous blood absorbs oxygen from and loses CO₂ to inspired gas. The path for airflow, in contrast, reaches a dead end at the alveolar walls; thus the alveolar space must be ventilated tidally, with inflow of fresh gas and outflow of alveolar gas alternating periodically at the respiratory rate (RR). To provide an enormous alveolar surface area (typically 70 m²) for blood-gas diffusion within the modest volume of a thoracic cavity (typically 7 L), nature has distributed both blood flow and ventilation among millions of tiny alveoli through multigenerational branching of both pulmonary arteries and bronchial airways. As a consequence of variations in tube lengths and calibers along these pathways as well as the effects of gravity, tidal pressure fluctuations, and anatomic constraints from the chest wall, the alveoli vary in their relative ventilations and perfusions. Not surprisingly, for the lung to be most efficient in exchanging gas, the fresh gas ventilation of a given alveolus must be matched to its perfusion.

For the respiratory system to succeed in oxygenating blood and eliminating CO₂, it must be able to ventilate the lung tidally and thus to freshen alveolar gas; it must provide for perfusion of the individual alveolus in a manner proportional to its ventilation; and it must allow adequate diffusion of respiratory gases between alveolar gas and capillary blood. Furthermore, it must accommodate several-fold increases in the demand for oxygen uptake or CO₂ elimination imposed by metabolic needs or acid-base derangement. Given these multiple requirements for normal operation, it is not surprising that many diseases disturb respiratory function. This chapter considers in some detail the physiologic determinants of lung ventilation and perfusion, elucidates how the matching distributions of these processes and rapid gas diffusion allow normal gas exchange, and discusses how common diseases derange these normal functions, thereby impairing gas exchange—or at least increasing the work required by the respiratory muscles or heart to maintain adequate respiratory function.

VENTILATION

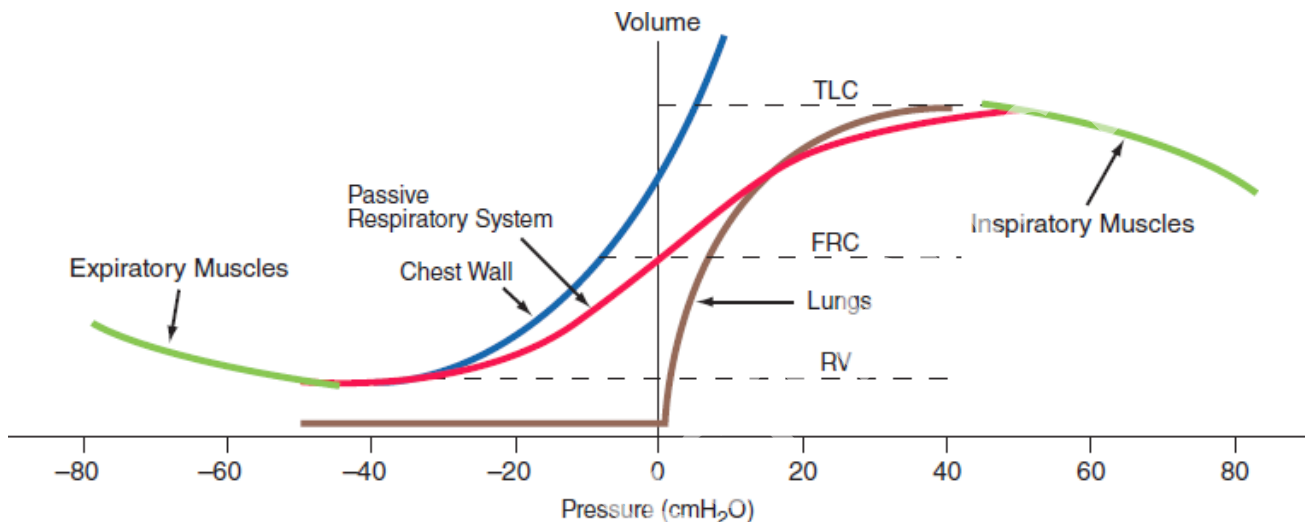
It is useful to conceptualize the respiratory system as three independently functioning components: the lung, including its airways; the neuromuscular system; and the chest wall, which includes everything that is not lung or active neuromuscular system. Accordingly, the mass of the respiratory muscles is part of the chest wall, while the force these muscles generate is part of the neuromuscular system; the abdomen (especially an obese abdomen) and the heart (especially an enlarged heart) are, for these purposes, part of the chest wall. Each of these three components has mechanical properties that relate to its enclosed volume (or—in the case of the neuromuscular system—the respiratory system volume at which it is operating) and to the rate of change of its volume (i.e., flow).

Volume-Related Mechanical Properties—Statics

Figure 279-1 shows the volume-related properties of each component of the respiratory system. Because of both surface tension at the air-liquid interface between alveolar wall lining fluid and alveolar gas and elastic recoil of the lung tissue itself, the lung requires a positive transmural pressure difference between alveolar gas and its pleural surface to stay inflated; this difference is called the *elastic recoil pressure* of the lung, and it increases with lung volume. The lung becomes rather stiff at high volumes, so that relatively small volume changes are accompanied by large changes in transpulmonary pressure; in contrast, the lung is compliant at lower volumes, including those at which tidal breathing normally occurs. At zero inflation pressure, even normal lungs retain some air in the alveoli. Because the small peripheral airways are tethered open by outward radial pull from inflated lung parenchyma attached to adventitia, as the lung deflates during exhalation, those small airways are pulled open progressively less, and eventually close, trapping some gas in the alveoli. This effect can be exaggerated with age and especially with obstructive airway diseases, resulting in gas trapping at quite large lung volumes.

FIGURE 279-1

Pressure-volume curves of the isolated lung, isolated chest wall, combined respiratory system, inspiratory muscles, and expiratory muscles. FRC, functional residual capacity; RV, residual volume; TLC, total lung capacity.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

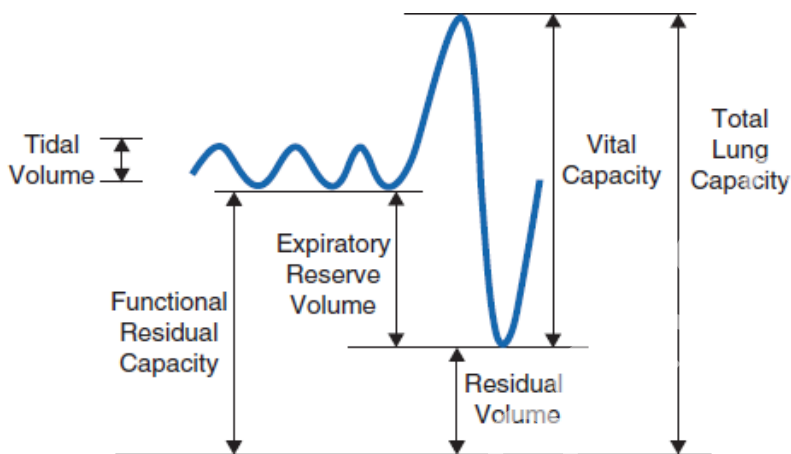
The elastic behavior of the passive chest wall (i.e., in the absence of neuromuscular activation) differs markedly from that of the lung. Whereas the lung tends toward full deflation with no distending (transmural) pressure, the chest wall encloses a large volume when pleural pressure equals body surface (atmospheric) pressure. Furthermore, the chest wall is compliant at high enclosed volumes, readily expanding even further in response to increases in transmural pressure. The chest wall also remains compliant at small negative transmural pressures (i.e., when pleural pressure falls slightly below atmospheric pressure), but as the volume enclosed by the chest wall becomes quite small in response to large negative transmural pressures, the passive chest wall becomes stiff due to squeezing together of ribs and intercostal muscles, diaphragm stretch, displacement of abdominal contents, and straining of ligaments and bony articulations. Under normal circumstances, the lung and the passive chest wall enclose essentially the same volume, the only difference being the volumes of the pleural fluid and of the lung parenchyma (normally both quite small in the absence of disease). For this reason and because the lung and chest wall function in mechanical series, the pressure required to displace the passive respiratory system (lungs plus chest wall) at any volume is simply the sum of the elastic recoil pressure of the lungs and the transmural pressure across the chest wall. When plotted against respiratory system volume, this relationship assumes a sigmoid shape, exhibiting stiffness at high lung volumes (imparted by the lung), stiffness at low lung volumes (imparted by the chest wall or sometimes by airway closure), and compliance in the middle range of lung volumes where normal tidal breathing occurs. In addition, a passive resting point of the respiratory system is attained when alveolar gas pressure equals body surface pressure (i.e., when the transrespiratory system pressure is zero). At this volume (called the *functional residual capacity* [FRC]), the outward recoil of the chest wall is balanced

exactly by the inward recoil of the lung. As these recoils are transmitted through the pleural fluid, the lung is pulled both outward and inward simultaneously at FRC, and thus its pressure falls below atmospheric pressure (typically, $-5 \text{ cmH}_2\text{O}$).

The normal passive respiratory system would equilibrate at the FRC and remain there were it not for the actions of the respiratory muscles. The inspiratory muscles act on the chest wall to generate the equivalent of positive pressure across the lungs and passive chest wall, while the expiratory muscles generate the equivalent of negative transrespiratory pressure. The maximal pressures these sets of muscles can generate vary with the lung volume at which they operate. This variation is due to length-tension relationships in striated muscle sarcomeres and to changes in mechanical advantage as the angles of insertion change with lung volume (Fig. 279-1). Nonetheless, under normal conditions, the respiratory muscles are substantially “overpowered” for their roles and generate more than adequate force to drive the respiratory system to its stiffness extremes, as determined by the lung (total lung capacity [TLC]) or by chest wall or airway closure (residual volume [RV]); the airway closure always prevents the adult lung from emptying completely under normal circumstances. The excursion between full and minimal lung inflation is called *vital capacity* (VC; Fig. 279-2) and is readily seen to be the difference between volumes at two unrelated stiffness extremes—one determined by the lung (TLC) and the other by the chest wall or airways (RV). Thus, although VC is easy to measure (see below), it provides little information about the intrinsic properties of the respiratory system. As will become clear, it is much more useful for the clinician to consider TLC and RV individually.

FIGURE 279-2

Spirogram demonstrating a slow vital capacity maneuver and various lung volumes.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Flow-Related Mechanical Properties—Dynamics

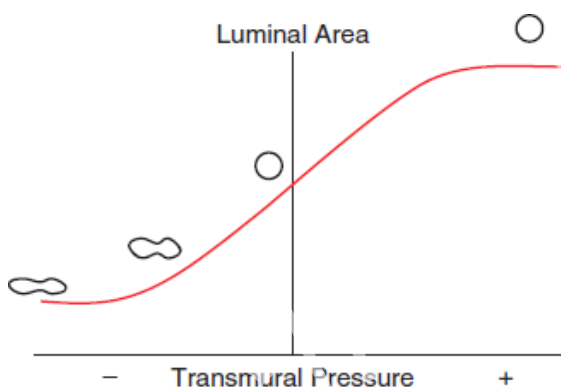
The passive chest wall and active neuromuscular system both exhibit mechanical behaviors related to the rate of change of volume, but these behaviors become quantitatively important only at markedly supraphysiologic breathing frequencies (e.g., during high-frequency mechanical ventilation), and thus will not be addressed here. In contrast, the dynamic airflow properties of the lung substantially affect its ability to ventilate and contribute importantly to the work of breathing, and these properties are often deranged by disease. Understanding dynamic airflow properties is, therefore, worthwhile.

As with the flow of any fluid (gas or liquid) in any tube, maintenance of airflow within the pulmonary airways requires a pressure gradient that falls along the direction of flow, the magnitude of which is determined by the flow rate and the frictional resistance to flow. During quiet tidal breathing, the pressure gradients driving inspiratory or expiratory flow are small owing to the very low frictional resistance of normal pulmonary airways (R_{aw} , normally $<2 \text{ cmH}_2\text{O/L/s}$). However, during rapid exhalation, another phenomenon reduces flow below that which would have been expected if frictional resistance were the only impediment to flow. This phenomenon is called *dynamic airflow limitation*, and it occurs because the bronchial airways through which air is

exhaled are collapsible rather than rigid (Fig. 279-3). An important anatomic feature of the pulmonary airways is its treelike branching structure. While the individual airways in each successive generation, from most proximal (trachea) to most distal (respiratory bronchioles), are smaller than those of the parent generation, their number increases exponentially such that the summed cross-sectional area of the airways becomes very large toward the lung periphery. Because flow (volume/time) is constant along the airway tree, the velocity of airflow (flow/summed cross-sectional area) is much greater in the central airways than in the peripheral airways. During exhalation, gas leaving the alveoli must, therefore, gain velocity as it proceeds toward the mouth. The energy required for this “convective” acceleration is drawn from the component of gas energy manifested as its local pressure, which reduces intraluminal gas pressure, airway transmural pressure, airway size (Fig. 279-3), and flow. This phenomenon is the Bernoulli effect, the same effect that keeps an airplane airborne, generating a lifting force by decreasing pressure above the curved upper surface of the wing due to acceleration of air flowing over the wing. If an individual tries to exhale more forcefully, the local velocity increases further and reduces airway size further, resulting in no net increase in flow. Under these circumstances, flow has reached its maximum possible value, or its *flow limit*. Lungs normally exhibit such dynamic airflow limitation. This limitation can be assessed by spirometry, in which an individual inhales fully to TLC and then forcibly exhales to RV. One useful spirometric measure is the volume of air exhaled during the forced expiratory volume 1 s (FEV₁), as discussed later. Maximal expiratory flow at any lung volume is determined by gas density, airway cross-section and distensibility, elastic recoil pressure of the lung, and frictional pressure loss to the flow-limiting airway site. Under normal conditions, maximal expiratory flow falls with lung volume (Fig. 279-4), primarily because of the dependence of lung recoil pressure on lung volume (Fig. 279-1). In pulmonary fibrosis, lung recoil pressure is increased at any lung volume, and thus the maximal expiratory flow is elevated when considered in relation to lung volume. Conversely, in emphysema, lung recoil pressure is reduced; this reduction is a principal mechanism by which maximal expiratory flows fall. Diseases that narrow the airway lumen at any transmural pressure (e.g., asthma or chronic bronchitis) or that cause excessive airway collapsibility (e.g., tracheomalacia) also reduce maximal expiratory flow.

FIGURE 279-3

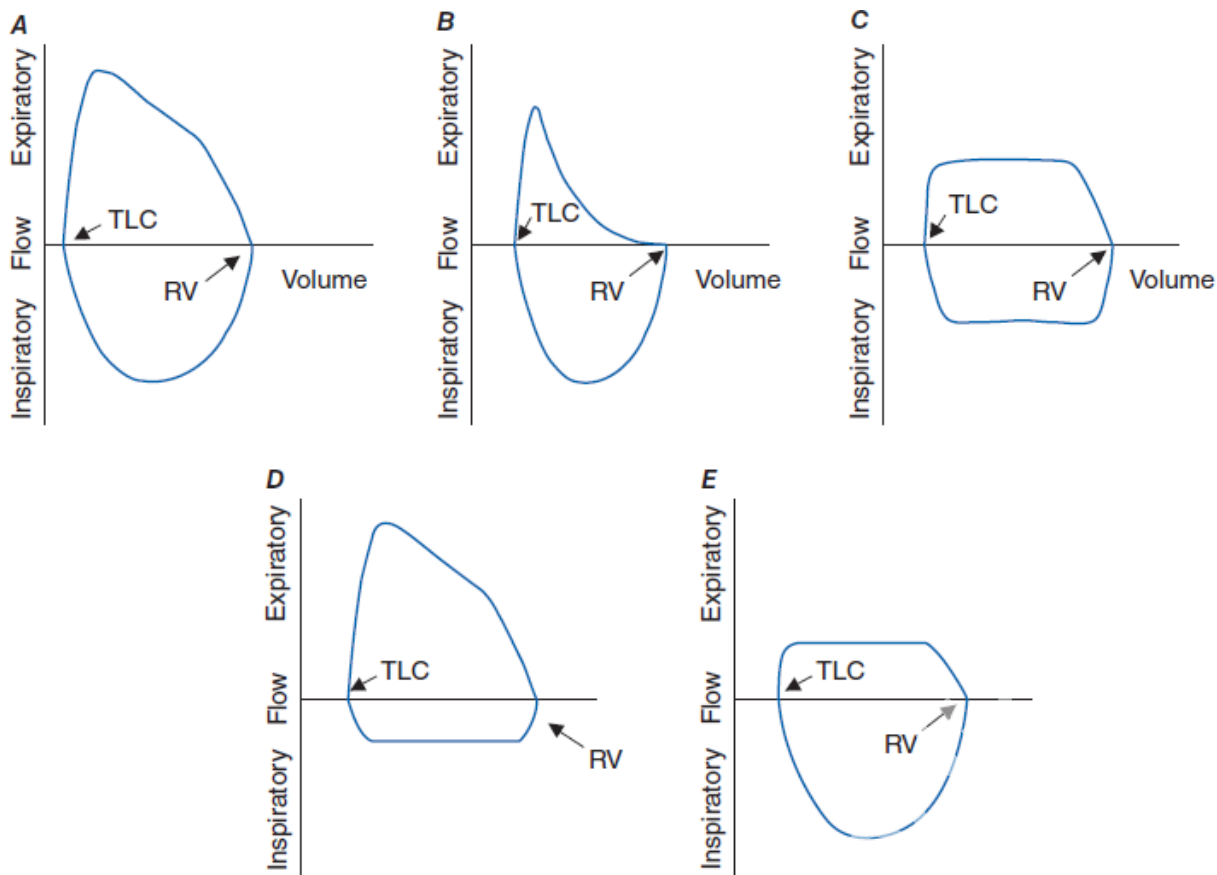
Luminal area versus transmural pressure relationship. Transmural pressure represents the pressure difference across the airway wall from inside to outside.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 279-4

Flow-volume loops. **A.** Normal. **B.** Airflow obstruction. **C.** Fixed central airway obstruction (either above or below the thoracic inlet). **D.** Variable upper airway obstruction (above the thoracic inlet) **E.** Variable lower airway obstruction (below the thoracic inlet); RV, residual volume; TLC, total lung capacity.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The Bernoulli effect also applies during inspiration, but the more negative pleural pressures during inspiration lower the pressure outside of the airways, thereby increasing transmural pressure and promoting airway expansion. Thus, inspiratory airflow limitation seldom occurs due to diffuse pulmonary airway disease. Conversely, extrathoracic airway narrowing (e.g., due to a tracheal adenoma or post-tracheostomy stricture) can lead to inspiratory airflow limitation (Fig. 279-4).

The Work of Breathing

In health, the elastic (volume change-related) and dynamic (flow-related) loads that must be overcome to ventilate the lungs at rest are small, and the work required of the respiratory muscles is minimal. However, the work of breathing can increase considerably due to a metabolic requirement for substantially increased ventilation, an abnormally increased mechanical load, or both. As discussed below, the rate of ventilation is primarily set by the need to eliminate carbon dioxide, and thus ventilation increases during exercise (sometimes by >20-fold) and during metabolic acidosis as a compensatory response. Naturally, the work rate required to overcome the elasticity of the respiratory system increases with both the depth and the frequency of tidal breaths, while the work required to overcome the dynamic load increases with total ventilation. A modest increase of ventilation is most efficiently achieved by increasing tidal volume but not RR, which is the normal ventilatory response to lower-level exercise. At higher levels of exercise, deep breathing persists, but RR also increases.

The work of breathing also increases when disease reduces the compliance of the respiratory system or increases the resistance to airflow. The former occurs commonly in diseases of the lung parenchyma (interstitial processes or fibrosis, alveolar filling diseases such as pulmonary edema or pneumonia, or substantial lung resection), and the latter occurs in obstructive airway diseases such as asthma, chronic bronchitis, emphysema, and cystic fibrosis. Furthermore, severe airflow obstruction can functionally reduce the compliance of the respiratory system by leading to dynamic hyperinflation. In this scenario, expiratory flows slowed by the obstructive airways disease may be insufficient to allow complete exhalation during the expiratory phase of tidal breathing; as a result, the "functional residual capacity (FRC)" from which the next breath is inhaled is greater than the

static FRC. With repetition of incomplete exhalations of each tidal breath, the operating FRC becomes dynamically elevated, sometimes to a level that approaches TLC. At these high lung volumes, the respiratory system is much less compliant than at normal breathing volumes, and thus the elastic work of each tidal breath is also increased. The dynamic pulmonary hyperinflation that accompanies severe airflow obstruction causes patients to sense difficulty in inhaling—even though the root cause of this pathophysiologic abnormality is expiratory airflow obstruction.

Adequacy of Ventilation

As noted above, the respiratory control system that sets the rate of ventilation responds to chemical signals, including arterial CO_2 and oxygen tensions and blood pH, and to volitional needs, such as the need to inhale deeply before playing a long phrase on the trumpet. Disturbances in ventilation are discussed in [Chap. 290](#). The focus of this chapter is on the relationship between ventilation of the lung and CO_2 elimination.

At the end of each tidal exhalation, the conducting airways are filled with alveolar gas that did not reach the mouth when expiratory flow stopped. During the ensuing inhalation, fresh gas immediately enters the airway tree at the mouth, but the gas first entering the alveoli at the start of inhalation is that same alveolar gas in the conducting airways that had just left the alveoli. Accordingly, fresh gas does not enter the alveoli until the volume of the conducting airways has been inspired. This volume is called the *anatomic dead space* (V_D). Quiet breathing with tidal volumes smaller than the anatomic dead space introduces no fresh gas into the alveoli at all; only that part of the inspired tidal volume (V_T) that is greater than the V_D introduces fresh gas into the alveoli. The dead space can be further increased functionally if some of the inspired tidal volume is delivered to a part of the lung that receives no pulmonary blood flow and thus cannot contribute to gas exchange (e.g., the portion of the lung distal to a large pulmonary embolus). In this situation, exhaled minute ventilation ($\dot{V}_E = V_T \times \text{RR}$) includes a component of dead space ventilation ($\dot{V}_D = V_D \times \text{RR}$) and a component of fresh gas alveolar ventilation ($\dot{V}_A = [V_T - V_D] \times \text{RR}$). CO_2 elimination from the alveoli is equal to \dot{V}_A times the difference in CO_2 fraction between inspired air (essentially zero) and alveolar gas (typically ~5.6% after correction for humidification of inspired air, corresponding to 40 mmHg). In the steady state, the alveolar fraction of CO_2 is equal to metabolic CO_2 production divided by alveolar ventilation. Because, as discussed below, alveolar and arterial CO_2 tensions are equal, and because the respiratory controller normally strives to maintain arterial P_{CO_2} (P_{aCO_2}) at ~40 mmHg, the adequacy of alveolar ventilation is reflected in P_{aCO_2} . If the P_{aCO_2} falls much below 40 mmHg, alveolar hyperventilation is present; if the P_{aCO_2} exceeds 40 mmHg, alveolar hypoventilation is present. Ventilatory failure is characterized by extreme alveolar hypoventilation.

As a consequence of oxygen uptake of alveolar gas into capillary blood, alveolar oxygen tension falls below that of inspired gas. The rate of oxygen uptake (determined by the body's metabolic oxygen consumption) is related to the average rate of metabolic CO_2 production, and their ratio—the “respiratory quotient” ($R = \dot{V}_{\text{CO}_2} / \dot{V}_{\text{O}_2}$)—depends largely on the fuel being metabolized. For a typical American diet, R is usually around 0.85. Together, these phenomena allow the estimation of alveolar oxygen tension, according to the following relationship, known as the *alveolar gas equation*:

$$P_{\text{aO}_2} = F_{\text{IO}_2} \times (P_{\text{bar}} - P_{\text{H}_2\text{O}}) - P_{\text{aCO}_2} / R$$

The alveolar gas equation also highlights the influences of inspired oxygen fraction (F_{IO_2}), barometric pressure (P_{bar}), and vapor pressure of water ($P_{\text{H}_2\text{O}} = 47$ mmHg at 37°C) in addition to alveolar ventilation (which sets P_{aCO_2}) in determining P_{aO_2} . An implication of the alveolar gas equation is that severe arterial hypoxemia rarely occurs as a pure consequence of alveolar hypoventilation at sea level while an individual is breathing air. The potential for alveolar hypoventilation to induce severe hypoxemia with otherwise normal lungs increases as P_{bar} falls with increasing altitude.

GAS EXCHANGE

Diffusion

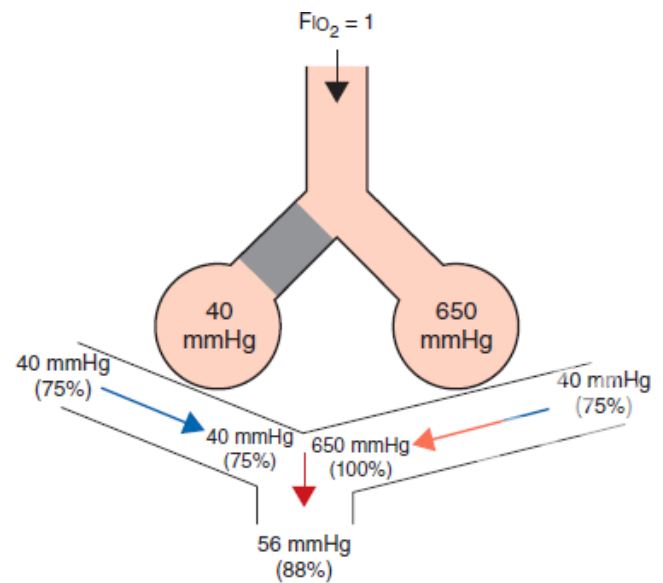
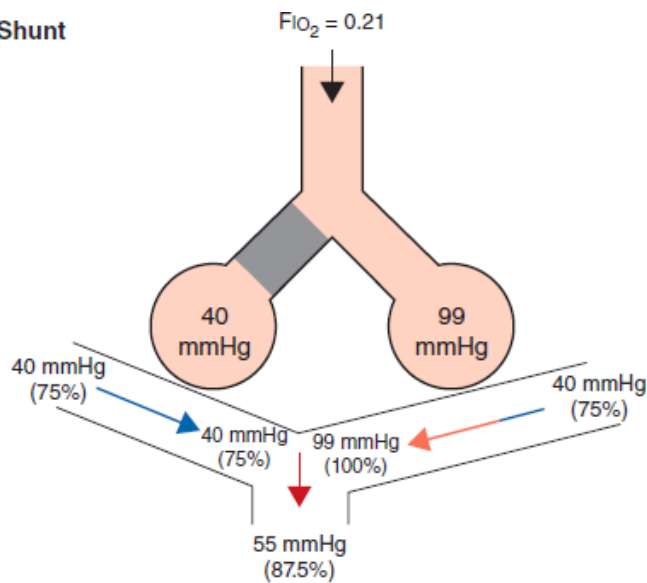
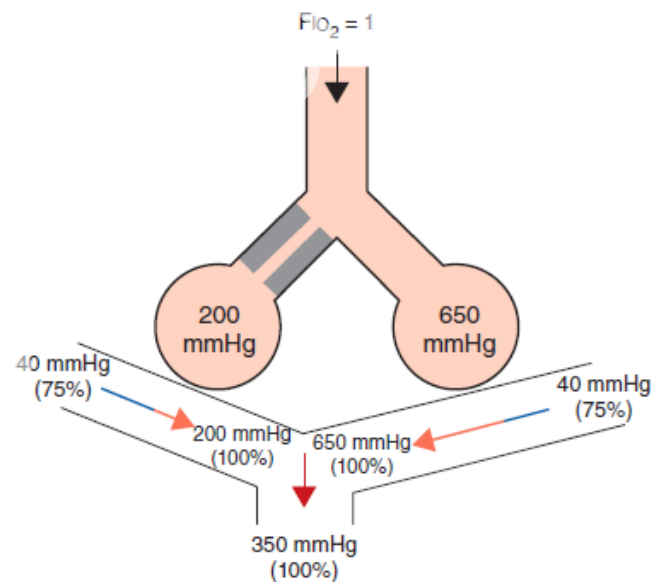
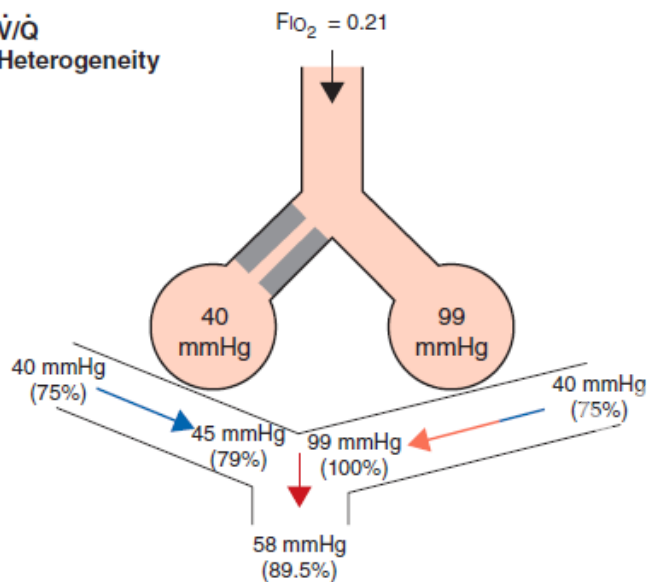
For oxygen to be delivered to the peripheral tissues, it must pass from alveolar gas into alveolar capillary blood by diffusing through alveolar membrane. The aggregate alveolar membrane is highly optimized for this process, with a very large surface area and minimal thickness. Diffusion through the alveolar membrane is so efficient in the human lung that in most circumstances hemoglobin of a red blood cell becomes fully oxygen saturated by the time the cell has traveled just one-third the length of the alveolar capillary. Thus, the uptake of alveolar oxygen is ordinarily limited by the amount of blood transiting the alveolar capillaries rather than by the rapidity with which oxygen can diffuse across the membrane; consequently, oxygen uptake from the lung is said to be “perfusion limited” rather than diffusion limited. CO₂ also equilibrates rapidly across the alveolar membrane. Therefore, the oxygen and CO₂ tensions in capillary blood leaving a normal alveolus are essentially equal to those in alveolar gas. Only in rare circumstances (e.g., at high altitude or in high-performance athletes exerting maximal effort) is oxygen uptake from normal lungs diffusion limited. Diffusion limitation can also occur in interstitial lung disease if substantially thickened alveolar walls remain perfused.

Ventilation/Perfusion Heterogeneity

As noted above, for gas exchange to be most efficient, ventilation to each individual alveolus (among the millions of alveoli) should match perfusion to its accompanying capillaries. Because of the differential effects of gravity on lung mechanics and blood flow throughout the lung and because of differences in airway and vascular architecture among various respiratory paths, there is minor ventilation/perfusion heterogeneity even in the normal lung; however, \dot{V}/\dot{Q} heterogeneity can be particularly marked in disease. Two extreme examples are (1) ventilation of unperfused lung distal to a pulmonary embolus, in which ventilation of the physiologic dead space is “wasted” in the sense that it does not contribute to gas exchange; and (2) perfusion of nonventilated lung (a “shunt”), which allows venous blood to pass through the lung unaltered. When mixed with fully oxygenated blood leaving other well-ventilated lung units, shunted venous blood disproportionately lowers the mixed arterial PaO₂ as a result of the nonlinear oxygen content versus Po₂ relationship of hemoglobin (**Fig. 279-5**). Furthermore, the resulting arterial hypoxemia is refractory to supplemental inspired oxygen. The reason is that (1) raising the inspired FI_{O₂} has no effect on alveolar gas tensions in nonventilated alveoli and (2) while raising inspired FI_{O₂} increases Pa_{CO₂} in ventilated alveoli, the oxygen content of blood exiting ventilated units increases only slightly, as hemoglobin will already have been nearly fully saturated and the solubility of oxygen in plasma is quite small.

FIGURE 279-5

Influence of air versus oxygen breathing on mixed arterial oxygenation in shunt and ventilation/perfusion heterogeneity. Partial pressure of oxygen (mmHg) and oxygen saturations are shown for mixed venous blood, for end capillary blood (normal vs affected alveoli), and for mixed arterial blood. FI_{O₂}, fraction of inspired oxygen; \dot{V}/\dot{Q} , ventilation/perfusion.

Shunt **\dot{V}/\dot{Q} Heterogeneity**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

A more common occurrence than the two extreme examples given above is a widening of the distribution of ventilation/perfusion ratios; such \dot{V}/\dot{Q} heterogeneity is a common consequence of lung disease. In this circumstance, perfusion of relatively underventilated alveoli results in the incomplete oxygenation of exiting blood. When mixed with well-oxygenated blood leaving higher \dot{V}/\dot{Q} regions, this partially reoxygenated blood disproportionately lowers arterial Pa_{O_2} , although to a lesser extent than does a similar perfusion fraction of blood leaving regions of pure shunt. In addition, in contrast to shunt regions, inhalation of supplemental oxygen raises the Pa_{O_2} , even in relatively underventilated low \dot{V}/\dot{Q} regions, and so the arterial hypoxemia induced by \dot{V}/\dot{Q} heterogeneity is typically responsive to oxygen therapy (Fig. 279-5).

In sum, arterial hypoxemia can be caused by substantial reduction of inspired oxygen tension, severe alveolar hypoventilation, perfusion of relatively underventilated (low \dot{V}/\dot{Q}) or completely unventilated (shunt) lung regions, and, in very unusual circumstances, by limitation of gas diffusion.

PATHOPHYSIOLOGY

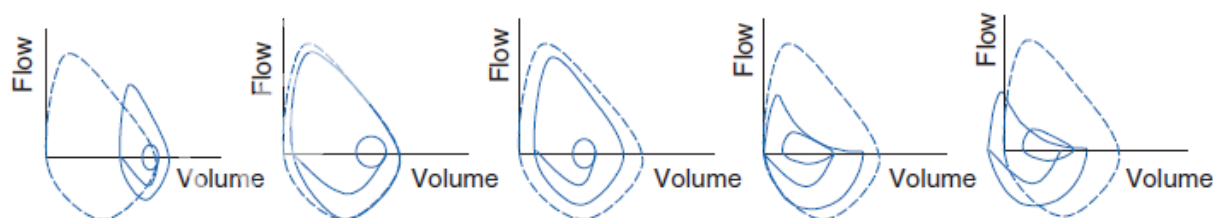
Although many diseases injure the respiratory system, this system responds to injury in relatively few ways. For this reason, the pattern of physiologic abnormalities may or may not provide sufficient information by which to discriminate among conditions.

Figure 279-6 lists abnormalities in pulmonary function testing that are typically found in a number of common respiratory disorders and highlight the simultaneous occurrence of multiple physiologic abnormalities. The coexistence of some of these respiratory disorders results in more complex superposition of these abnormalities. Methods to measure respiratory system function clinically are described later in this chapter.

FIGURE 279-6

Common abnormalities of pulmonary function (see text). Pulmonary function values are expressed as a percentage of normal predicted values, except for R_{aw} , which is expressed as cmH₂O/L/s (normal, <2 cmH₂O/L/s). The figures at the bottom of each column show the typical configuration of flow-volume loops in each condition, including the flow-volume relationship during tidal breathing. b.d., bronchodilator; DL_{CO}, diffusion capacity of lung for carbon monoxide; FEV₁, forced expiratory volume in 1 s; FRC, functional residual capacity; FVC, forced vital capacity; R_{aw} , airways resistance; RV, residual volume; TLC, total lung capacity.

	Restriction due to increased lung elastic recoil (pulmonary fibrosis)	Restriction due to chest wall abnormality (moderate obesity)	Restriction due to respiratory muscle weakness (myasthenia gravis)	Obstruction due to airway narrowing (acute asthma)	Obstruction due to decreased elastic recoil (severe emphysema)
TLC	60%	95%	75%	100%	130%
FRC	60%	65%	100%	104%	220%
RV	60%	100%	120%	120%	310%
FVC	60%	92%	60%	90%	60%
FEV ₁	75%	92%	60%	35% pre-b.d. 75% post-b.d.	35% pre-b.d. 38% post-b.d.
R_{aw}	1.0	1.0	1.0	2.5	1.5
DL _{CO}	60%	95%	80%	120%	40%



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Ventilatory Restriction due to Increased Elastic Recoil—Example: Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis raises lung recoil at all lung volumes, thereby lowering TLC, FRC, and RV as well as forced vital capacity (FVC). Maximal expiratory flows are also reduced from normal values but are elevated when considered in relation to lung volumes. Increased flow occurs both because the increased lung recoil drives greater maximal flow at any lung volume and because airway diameters are relatively increased due to greater radially outward traction exerted on bronchi by the stiff lung parenchyma. For the same reason, airway resistance is also normal. Destruction of the pulmonary capillaries by the fibrotic process results in a marked reduction in diffusing capacity (see below). Oxygenation is often severely reduced by persistent perfusion of alveolar units that are relatively underventilated due to fibrosis of nearby (and mechanically linked) lung. The flow-

volume loop (see below) looks like a miniature version of a normal loop but is shifted toward lower absolute lung volumes and displays maximal expiratory flows that are increased for any given volume over the normal tracing.

Ventilatory Restriction due to Chest Wall Abnormality—Example: Moderate Obesity

As the size of the average American continues to increase, this pattern may become the most common of pulmonary function abnormalities. In moderate obesity, the outward recoil of the chest wall is blunted by the weight of chest wall fat and the space occupied by intra-abdominal fat. In this situation, preserved inward recoil of the lung overbalances the reduced outward recoil of the chest wall, and FRC falls. Because respiratory muscle strength and lung recoil remain normal, TLC is typically unchanged (although it may fall in massive obesity) and RV is normal (but may be reduced in massive obesity). Mild hypoxemia may be present due to perfusion of alveolar units that are poorly ventilated because of airway closure in dependent portions of the lung during breathing near the reduced FRC. Flows remain normal, as does the diffusion capacity of the lung for carbon monoxide (DL_{CO}), unless obstructive sleep apnea (which often accompanies obesity) and associated chronic intermittent hypoxemia have induced pulmonary arterial hypertension, in which case DL_{CO} may be low.

Ventilatory Restriction due to Reduced Muscle Strength—Example: Myasthenia Gravis

In this circumstance, FRC remains normal, as both lung recoil and passive chest wall recoil are normal. However, TLC is low and RV is elevated because respiratory muscle strength is insufficient to push the passive respiratory system fully toward either volume extreme. Caught between the low TLC and the elevated RV, FVC, and FEV_1 are reduced as “innocent bystanders.” As airway size and lung vasculature are unaffected, both R_{aw} and DL_{CO} are normal. Oxygenation is normal unless weakness becomes so severe that the patient has insufficient strength to reopen collapsed alveoli during sighs, with resulting atelectasis.

Airflow Obstruction due to Decreased Airway Diameter—Example: Acute Asthma

During an episode of acute asthma, luminal narrowing due to smooth muscle constriction as well as inflammation and thickening within the small- and medium-sized bronchi raise frictional resistance and reduce airflow. “Scooping” of the flow-volume loop is caused by reduction of airflow, especially at lower lung volumes. Often, airflow obstruction can be reversed by inhalation of β_2 -adrenergic agonists acutely or by treatment with inhaled steroids chronically. TLC usually remains normal (although elevated TLC is sometimes seen in long-standing asthma), but FRC may be dynamically elevated. RV is often increased due to exaggerated airway closure at low lung volumes, and this elevation of RV reduces FVC. Because central airways are narrowed, R_{aw} is usually elevated. Mild arterial hypoxemia is often present due to perfusion of relatively underventilated alveoli distal to obstructed airways (and is responsive to oxygen supplementation), but DL_{CO} is normal or mildly elevated.

Airflow Obstruction due to Decreased Elastic Recoil—Example: Severe Emphysema

Loss of lung elastic recoil in severe emphysema results in pulmonary hyperinflation, of which elevated TLC is the hallmark. FRC is more severely elevated due to both loss of lung elastic recoil and dynamic hyperinflation—the same phenomenon as auto-PEEP (auto-positive end-expiratory pressure), which is the positive end-expiratory alveolar pressure that occurs when a new breath is initiated before the lung volume is allowed to return to FRC. RV is very severely elevated because of airway closure and because exhalation toward RV may take so long that RV cannot be reached before the patient must inhale again. Both FVC and FEV_1 are markedly decreased, the former because of the severe elevation of RV and the latter because loss of lung elastic recoil reduces the pressure driving maximal expiratory flow and also reduces tethering open of small intrapulmonary airways. The flow-volume loop demonstrates marked scooping, with an initial transient spike of flow attributable largely to expulsion of air from collapsing central airways at the onset of forced exhalation. Otherwise, the central airways remain relatively unaffected, so R_{aw} is normal in “pure” emphysema. Loss of alveolar surface and capillaries in the alveolar walls reduces DL_{CO} ; however, because poorly ventilated emphysematous acini are also poorly perfused (due to loss of their capillaries), arterial hypoxemia

usually is not seen at rest until emphysema becomes very severe. However, during exercise, Pa_{O_2} may fall precipitously if extensive destruction of the pulmonary vasculature prevents a sufficient increase in cardiac output and mixed venous oxygen content falls substantially. Under these circumstances, any venous admixture through low \dot{V}/\dot{Q} units has a particularly marked effect in lowering mixed arterial oxygen tension.

FUNCTIONAL MEASUREMENTS

Measurement of Ventilatory Function

LUNG VOLUMES

Figure 279-2 demonstrates a spirometry tracing in which the volume of air entering or exiting the lung is plotted over time. In a slow vital capacity maneuver, the patient inhales from FRC, fully inflating the lungs to TLC, and then exhales slowly to RV; VC, the difference between TLC and RV, represents the maximal excursion of the respiratory system. Spirometry discloses relative volume changes during these maneuvers but cannot reveal the absolute volumes at which they occur. To determine absolute lung volumes, two approaches are commonly used: inert gas dilution and body plethysmography. In the former, a known amount of a nonabsorbable inert gas (usually helium or neon) is inhaled in a single large breath or is rebreathed from a closed circuit; the inert gas is diluted by the gas resident in the lung at the time of inhalation, and its final concentration reveals the volume of resident gas contributing to the dilution. A drawback of this method is that regions of the lung that ventilate poorly (e.g., due to airflow obstruction) may not receive much inspired inert gas and so do not contribute to its dilution. Therefore, inert gas dilution (especially in the single-breath method) often underestimates true lung volumes.

In the second approach, FRC is determined by measuring the compressibility of gas within the chest, which is proportional to the volume of gas being compressed. The patient sits in a body plethysmograph (a chamber usually made of transparent plastic to minimize claustrophobia) and, at the end of a normal tidal breath (i.e., when lung volume is at FRC), is instructed to pant against a closed shutter, thus periodically compressing air within the lung slightly. Pressure fluctuations at the mouth and volume fluctuations within the body box (equal but opposite to those in the chest) are determined, and from these measurements, the thoracic gas volume is calculated by means of Boyle's law. Once FRC is obtained, TLC and RV are calculated by adding the value for inspiratory capacity and subtracting the value for expiratory reserve volume, respectively (both values having been obtained during spirometry) (Fig. 279-2). The most important determinants of healthy individuals' lung volumes are height, age, and sex, but there is considerable additional normal variation beyond that accounted for by these parameters. In addition, race influences lung volumes; on average, TLC values are ~12% lower in African Americans and 6% lower in Asian Americans than in Caucasian Americans. In practice, a mean "normal" value is predicted by multivariate regression equations using height, age, and sex, and the patient's value is divided by the predicted value (often with "race correction" applied) to determine "percent predicted." For most measures of lung function, 85–115% of the predicted value can be normal; however, in health, the various lung volumes tend to scale together. For example, if one is "normal big" with a TLC 110% of the predicted value, all other lung volumes and spirometry values will also approximate 110% of their respective predicted values. This pattern is particularly helpful in evaluating airflow, as discussed below.

AIR FLOW

As noted above, spirometry plays a key role in lung volume determination. Even more often, spirometry is used to measure airflow, which reflects the dynamic properties of the lung. During an FVC maneuver, the patient inhales to TLC and then exhales rapidly and forcefully to RV; this method ensures that flow limitation has been achieved, so that the precise effort made has little influence on actual flow. The total amount of air exhaled is the FVC, and the amount of air exhaled in the first second is the FEV_1 ; the FEV_1 is a flow rate, revealing volume change per time. Like lung volumes, an individual's maximal expiratory flows should be compared with predicted values based on height, age, and sex. While the FEV_1/FVC ratio is typically reduced in airflow obstruction, this condition can also reduce FVC by raising RV, sometimes rendering the FEV_1/FVC ratio "artificially normal"

with the erroneous implication that airflow obstruction is absent. To circumvent this problem, it is useful to compare FEV_1 as a fraction of its predicted value with TLC as a fraction of its predicted value. In health, the results are usually similar. In contrast, even an FEV_1 value that is 95% of its predicted value may actually be relatively low if TLC is 110% of its respective predicted value. In this case, airflow obstruction may be present, despite the “normal” value for FEV_1 .

The relationships among volume, flow, and time during spirometry are best displayed in two plots—the spirogram (volume vs time) and the flow-volume loop (flow vs volume) (Fig. 279-4). In conditions that cause airflow obstruction, the site of obstruction is sometimes correlated with the shape of the flow-volume loop. In diseases that cause lower airway obstruction, such as asthma and emphysema, flows decrease more rapidly with declining lung volumes, leading to a characteristic scooping of the flow-volume loop. In contrast, fixed upper-airway obstruction typically leads to inspiratory and/or expiratory flow plateaus (Fig. 279-4).

AIRWAYS RESISTANCE

The total resistance of the pulmonary and upper airways is measured in the same body plethysmograph used to measure FRC. The patient is asked once again to pant, but this time against a closed and then opened shutter. Panting against the closed shutter reveals the thoracic gas volume as described above. When the shutter is opened, flow is directed to and from the body box, so that volume fluctuations in the box reveal the extent of thoracic gas compression, which in turn reveals the pressure fluctuations driving flow. Simultaneous measurement of flow allows the calculation of lung resistance (as flow divided by pressure). In health, R_{aw} is very low (<2 cmH₂O/L/s), and half of the detected resistance resides within the upper airway. In the lung, most resistance originates in the central airways. For this reason, airways resistance measurement tends to be insensitive to peripheral airflow obstruction.

RESPIRATORY MUSCLE STRENGTH

To measure respiratory muscle strength, the patient is instructed to exhale or inhale with maximal effort against a closed shutter while pressure is monitored at the mouth. Pressures $>\pm 60$ cmH₂O at FRC are considered adequate and make it unlikely that respiratory muscle weakness accounts for any other resting ventilatory dysfunction that is identified.

Measurement of Gas Exchange

DIFFUSING CAPACITY (DL_{CO})

This test uses a small (and safe) amount of carbon monoxide (CO) to measure gas exchange across the alveolar membrane during a 10-s breath hold. CO in exhaled breath is analyzed to determine the quantity of CO crossing the alveolar membrane and combining with hemoglobin in red blood cells. This “single-breath diffusing capacity” (DL_{CO}), value increases with the surface area available for diffusion and the amount of hemoglobin within the capillaries, and it varies inversely with alveolar membrane thickness. Thus, DL_{CO} decreases in diseases that thicken or destroy alveolar membranes (e.g., pulmonary fibrosis, emphysema), curtail the pulmonary vasculature (e.g., pulmonary hypertension), or reduce alveolar capillary hemoglobin (e.g., anemia). Single-breath diffusing capacity may be elevated in acute congestive heart failure, asthma, polycythemia, and pulmonary hemorrhage.

Arterial Blood Gases

The effectiveness of gas exchange can be assessed by measuring the partial pressures of oxygen and CO₂ in a sample of blood obtained by arterial puncture. The oxygen content of blood (Ca_{O_2}) depends on arterial saturation (%O₂Sat), which is set by Pa_{O_2} , pH, and Pa_{CO_2} according to the oxyhemoglobin dissociation curve. Ca_{O_2} can also be measured by oximetry (see below):

$$\text{CaO}_2 \text{ (mL/dL)} = 1.39 \text{ (mL/dL)} \times [\text{hemoglobin}](\text{g}) \times \% \text{O}_2 \text{ Sat} + 0.003 \text{ (mL/dL/mmHg)} \times \text{PaO}_2 \text{ (mmHg)}$$

If hemoglobin saturation alone needs to be determined, this task can be accomplished noninvasively with pulse oximetry.

FURTHER READING

Bates JH: Systems physiology of the airways in health and obstructive pulmonary disease. *Wiley Interdiscip Rev Sys Biol Med* 8:423, 2016.

Hughes JM et al: Effect of lung volume on the distribution of pulmonary blood flow in man. *Respir Physiol* 4:58, 1968. [\[PubMed: 5639524\]](#)

Levitsky MG et al: *Pulmonary Physiology*, 8th ed. New York, McGraw Hill, 2013; <http://accessmedicine.mhmedical.com.ezproxy.uaeu.ac.ae/book.aspx?bookid=575>. Accessed June 6, 2017.

Macklem PT, Murphy BR: The forces applied to the lung in health and disease. *Am J Med* 57:371, 1974. [\[PubMed: 4416277\]](#)

Pederson OF, Ingram RH: Configuration of maximal expiratory flow-volume curve: Model experiments with physiologic implications. *J Appl Physiol* 58:1305, 1985. [\[PubMed: 3988684\]](#)

Prange HD: *Respiratory Physiology: Understanding Gas Exchange*. New York, Chapman and Hill, 1996.

Weibel ER: Morphometric estimation of pulmonary diffusion capacity, I. Model and method. *Respir Physiol* 11:54, 1970. [\[PubMed: 4992513\]](#)

West JB: *Respiratory Physiology, The Essentials*, 9th ed. Philadelphia, Lippincott Williams & Wilkins, 2012.

Wiley Online Library: Comprehensive Physiology: The Respiratory System. Available from <http://www.comprehensivephysiology.com/WileyCDA/Section/id-420557.html>. Accessed August 12, 2016.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of Drs. Steven E. Weinberger and Irene M. Rosen to this chapter in previous editions.

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library
Silverchair

Chapter 303: Cellular and Molecular Biology of the Kidney

Alfred L. George; Eric G. Neilson

INTRODUCTION

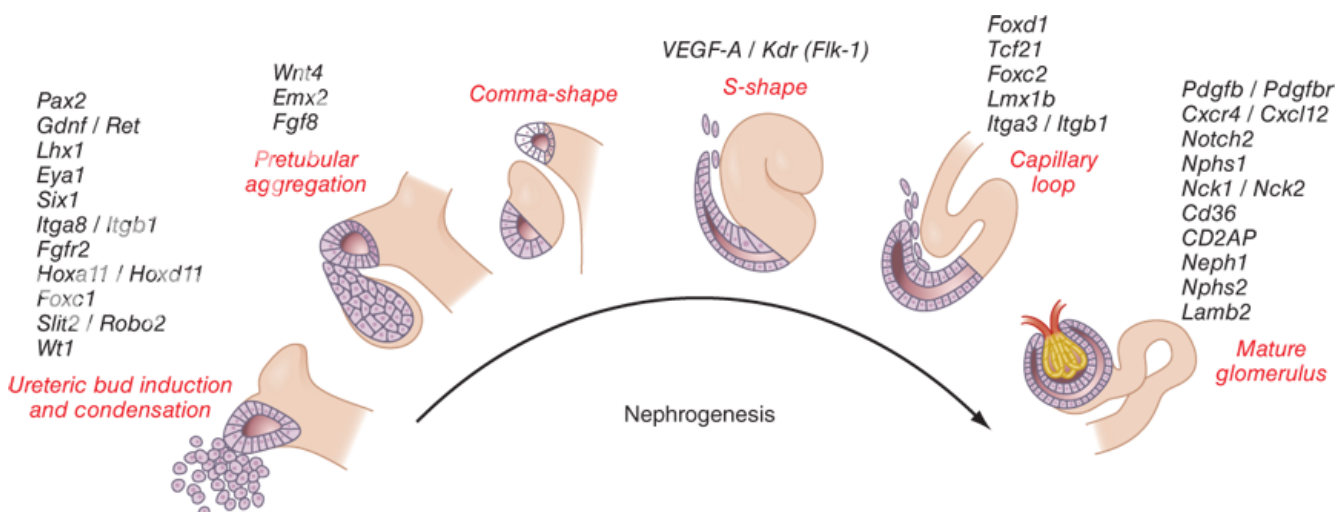
The kidney is one of the most highly differentiated organs in the body. At the conclusion of embryologic development, nearly 30 different cell types form a multitude of filtering capillaries and segmented nephrons enveloped by a dynamic interstitium. This cellular diversity modulates a variety of complex physiologic processes. Endocrine functions, the regulation of blood pressure and intraglomerular hemodynamics, solute and water transport, acid-base balance, and removal of drug metabolites are all accomplished by intricate mechanisms of renal response. This breadth of physiology hinges on the clever ingenuity of nephron architecture that evolved as complex organisms came out of water to live on land.

EMBRYOLOGIC DEVELOPMENT

Kidneys develop from intermediate mesoderm under the timed or sequential control of a growing number of genes, described in **Fig. 303-1**. The transcription of these genes is guided by morphogenic cues that invite two ureteric buds to each penetrate bilateral metanephric blastema, where they induce primary mesenchymal cells to form early nephrons. The two ureteric buds emerge from posterior nephric ducts and mature into separate collecting systems that eventually form a renal pelvis and ureter. Induced mesenchyme undergoes mesenchymal epithelial transitions to form comma-shaped bodies at the proximal end of each ureteric bud leading to the formation of S-shaped nephrons that cleft and enjoin with penetrating endothelial cells derived from sprouting angioblasts. Under the influence of vascular endothelial growth factor A (VEGF-A), these penetrating cells form capillaries with surrounding mesangial cells that differentiate into a glomerular filter for plasma water and solute. The ureteric buds branch and each branch produce a new set of nephrons. The number of branching events ultimately determines the total number of nephrons in each kidney. There are ~900,000 glomeruli in each kidney in normal birth weight adults and as few as 225,000 in low-birth-weight adults, with the latter producing numerous comorbid risks.

FIGURE 303-1

Genes controlling renal nephrogenesis. A growing number of genes have been identified at various stages of glomerulotubular development in the mammalian kidney. The genes listed have been tested in various genetically modified mice, and their location corresponds to the classical stages of kidney development postulated by Saxon in 1987.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Glomeruli evolve as complex capillary filters with fenestrated endothelia under the guiding influence of VEGF-A and angiopoietin-1 secreted by adjacently developing podocytes. Epithelial podocytes facing the urinary space envelop the exterior basement membrane supporting these emerging endothelial capillaries. Podocytes are partially polarized and periodically slough into the urinary space by epithelial-mesenchymal transition, and to a lesser extent apoptosis, only to be replenished by migrating parietal epithelia from Bowman capsule. Impaired replenishment results in heavy proteinuria. Podocytes attach to the basement membrane by special foot processes and share a slit-pore membrane with their neighbor. The slit-pore membrane forms a filter for plasma water and solute by the synthetic interaction of nephrin, annexin-4, CD2AP, FAT, ZO-1, P-cadherin, podocin, TRPC6, PLCE1, and Neph 1-3 proteins. Mutations in many of these proteins also result in heavy proteinuria. The glomerular capillaries are embedded in a mesangial matrix shrouded by parietal and proximal tubular epithelia forming Bowman capsule. Mesangial cells have an embryonic lineage consistent with arteriolar or juxtaglomerular cells and contain contractile actin-myosin fibers. These mesangial cells make contact with glomerular capillary loops, and their local matrix holds them in condensed arrangement.

Between nephrons lies the renal interstitium. This region forms a functional space surrounding glomeruli and their downstream tubules, which are home to resident and trafficking cells such as fibroblasts, dendritic cells, occasional lymphocytes, and lipid-laden macrophages. The cortical and medullary peritubular capillaries, which siphon off solute and water following tubular reclamation of glomerular filtrate, are also part of the interstitial fabric as well as a web of connective tissue that supports the kidney's emblematic architecture of folding tubules. The relational precision of these structures determines the unique physiology of the kidney.

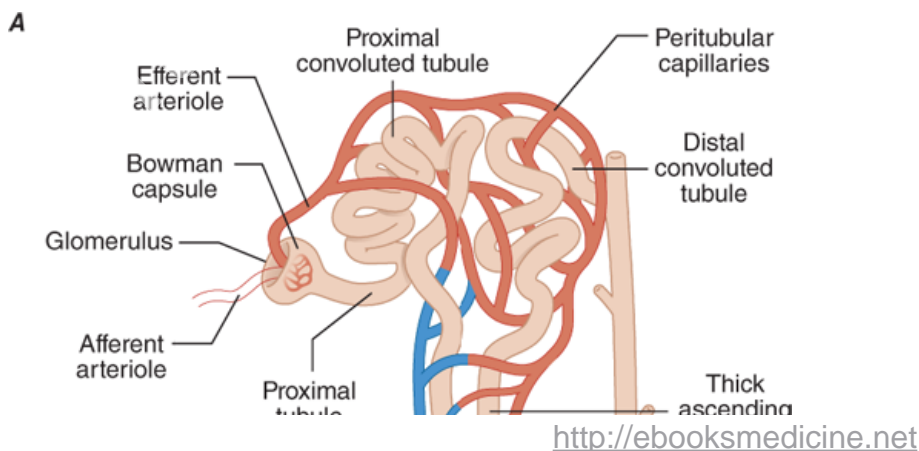
Each nephron is partitioned during embryologic development into a proximal tubule, descending and ascending limbs of the loop of Henle, distal tubule, and the collecting duct. These classic tubular segments build from subsegments lined by highly unique epithelia serving regional physiology. All nephrons have the same structural components, but there are two types whose structures depend on their location within the kidney. The majority of nephrons are cortical, with glomeruli located in the mid-to-outer cortex. Fewer nephrons are juxtamedullary, with glomeruli at the boundary of the cortex and outer medulla. Cortical nephrons have short loops of Henle, whereas juxtamedullary nephrons have long loops of Henle. There are critical differences in blood supply as well. The peritubular capillaries surrounding cortical nephrons are shared among adjacent nephrons. By contrast, juxtamedullary nephrons depend on individual capillaries called *vasa recta* that run alongside the long loops of Henle. Cortical nephrons perform most of the glomerular filtration because there are more of them and because their afferent arterioles are larger than their respective efferent arterioles. The juxtamedullary nephrons, with longer loops of Henle, create an osmotic gradient for concentrating urine. How developmental instructions specify the differentiation of all these unique epithelia among various tubular segments is still unknown.

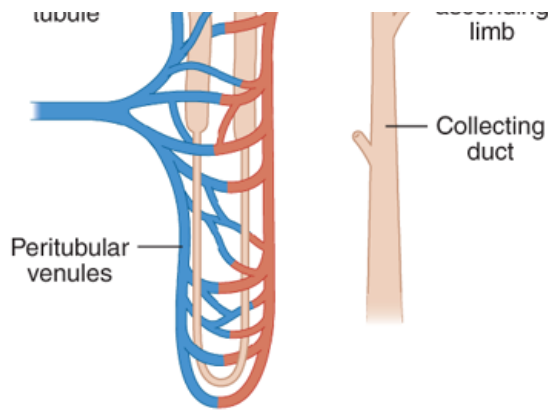
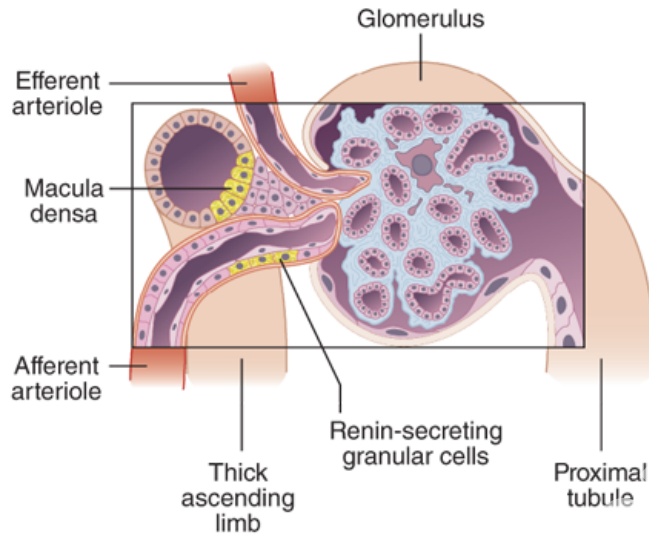
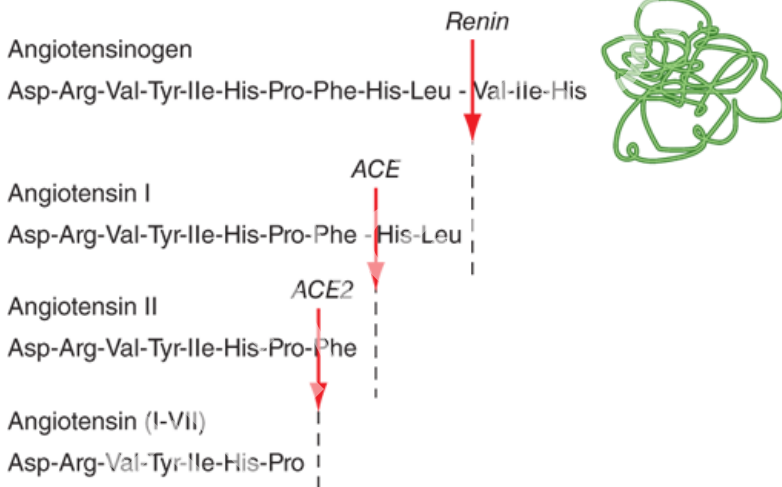
DETERMINANTS AND REGULATION OF GLOMERULAR FILTRATION

Renal blood flow normally drains ~20% of the cardiac output, or 1000 mL/min. Blood reaches each nephron through the afferent arteriole leading into a glomerular capillary where large amounts of fluid and solutes are filtered to form the tubular fluid. The distal ends of the glomerular capillaries coalesce to form an efferent arteriole leading to the first segment of a second capillary network (cortical peritubular capillaries or medullary vasa recta) surrounding the tubules (Fig. 303-2A). Thus, nephrons have two capillary beds arranged in a series separated by the efferent arteriole that regulates the hydrostatic pressure in both capillary beds. The distal capillaries empty into small venous branches that coalesce into larger veins to eventually form the renal vein.

FIGURE 303-2

Renal microcirculation and the renin-angiotensin system. A. Diagram illustrating relationships of the nephron with glomerular and peritubular capillaries. **B.** Expanded view of the glomerulus with its juxtaglomerular apparatus including the macula densa and adjacent afferent arteriole. **C.** Proteolytic processing steps in the generation of angiotensins.



**B****C**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The hydrostatic pressure gradient across the glomerular capillary wall is the primary driving force for glomerular filtration. Oncotic pressure within the capillary lumen, determined by the concentration of unfiltered plasma proteins, partially offsets the hydrostatic pressure gradient and opposes filtration. As the oncotic pressure rises along the length of the glomerular capillary, the driving force for filtration falls to zero en route to the efferent arteriole. Approximately 20% of the renal plasma flow is filtered into Bowman space, and the ratio of glomerular filtration rate (GFR) to renal plasma flow determines the filtration fraction. Several factors, mostly hemodynamic, contribute to the regulation of filtration under physiologic conditions.

Although glomerular filtration is affected by renal artery pressure, this relationship is not linear across the range of physiologic blood pressures due to autoregulation of GFR. Autoregulation of glomerular filtration is the result of three major factors that modulate either afferent or efferent arteriolar tone: these include an autonomous vasoreactive (myogenic) reflex in the afferent arteriole, *tubuloglomerular feedback* (TGF), and angiotensin II-mediated vasoconstriction of the efferent arteriole. The myogenic reflex is a first line of defense against fluctuations in renal blood flow. Acute changes in renal perfusion pressure evoke reflex constriction or dilatation of the afferent arteriole in response to increased or decreased pressure, respectively. This phenomenon helps protect the glomerular capillary from sudden changes in systolic pressure.

TGF changes the rate of filtration and tubular flow by reflex vasoconstriction or dilatation of the afferent arteriole. TGF is mediated by specialized cells in the thick ascending limb of the loop of Henle called the *macula densa* that act as sensors of solute concentration and tubular fluid flow rate. With high tubular flow rates, a proxy for an inappropriately high filtration rate, there is increased solute delivery to the macula densa (**Fig. 303-2B**) that evokes vasoconstriction of the afferent arteriole causing GFR to return toward normal. One component of the soluble signal from the macula densa is adenosine triphosphate (ATP) released by the cells during increased NaCl reabsorption. ATP is metabolized in the extracellular space to generate adenosine, a potent vasoconstrictor of the afferent arteriole. During conditions associated with a fall in filtration rate, reduced solute delivery to the macula densa attenuates TGF, allowing afferent arteriolar dilatation and restoring GFR to normal levels. Angiotensin II and reactive oxygen species enhance, while nitric oxide (NO) blunts TGF.

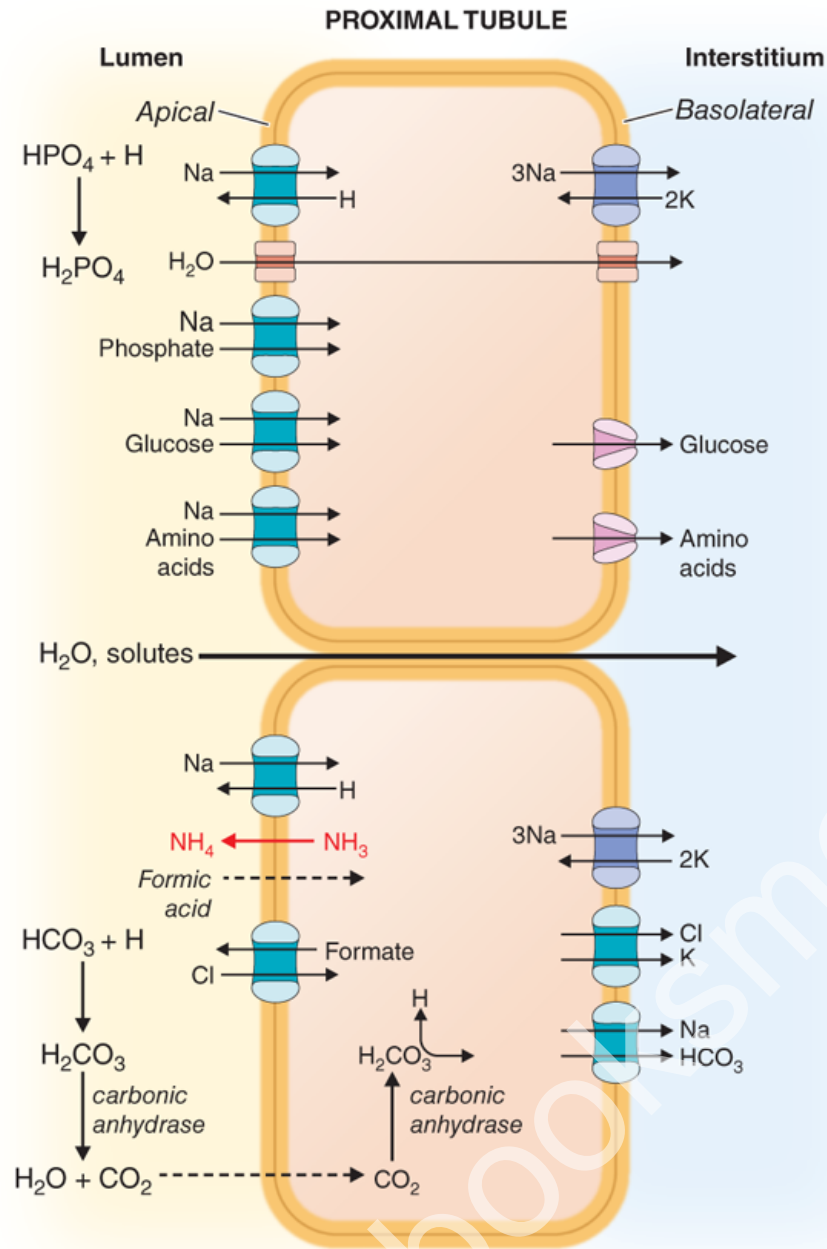
The third component underlying autoregulation of GFR involves angiotensin II. During states of reduced renal blood flow, renin is released from granular cells within the wall of the afferent arteriole near the macula densa in a region called the juxtaglomerular apparatus (**Fig. 303-2B**). Renin, a proteolytic enzyme, catalyzes the conversion of angiotensinogen to angiotensin I, which is subsequently converted to angiotensin II by angiotensin-converting enzyme (ACE) (**Fig. 303-2C**). Angiotensin II evokes vasoconstriction of the efferent arteriole, and the resulting increased glomerular hydrostatic pressure elevates GFR to normal levels.

MECHANISMS OF RENAL TUBULAR TRANSPORT

The renal tubules are composed of highly differentiated epithelia that vary dramatically in morphology and function along the nephron (**Fig. 303-3**). The cells lining the various tubular segments form monolayers connected to one another by a specialized region of the adjacent lateral membranes called the *tight junction*. Tight junctions form an occlusive barrier that separates the lumen of the tubule from the interstitial spaces surrounding the tubule and also apportions the cell membrane into discrete domains: the apical membrane facing the tubular lumen and the basolateral membrane facing the interstitium. This regionalization allows cells to allocate membrane proteins and lipids asymmetrically. Owing to this feature, renal epithelial cells are said to be *polarized*. The asymmetric assignment of membrane proteins, especially proteins mediating transport processes, provides the machinery for directional movement of fluid and solutes by the nephron.

FIGURE 303-3

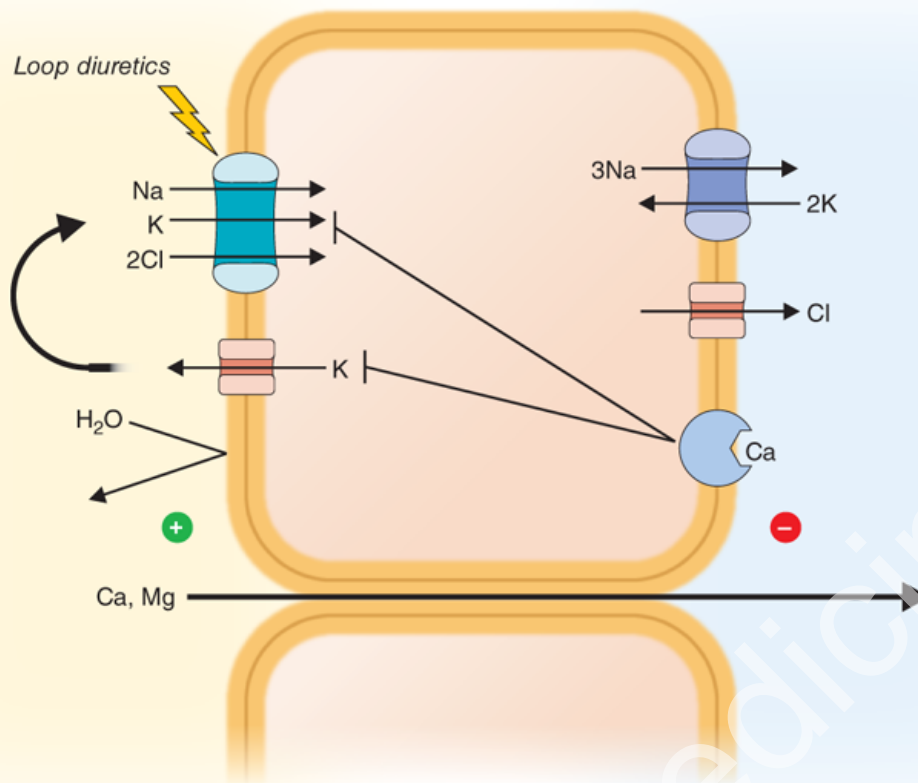
Transport activities of the major nephron segments. Representative cells from five major tubular segments are illustrated with the lumen side (apical membrane) facing left and interstitial side (basolateral membrane) facing right. **A.** Proximal tubular cells. **B.** Typical cell in the thick ascending limb of the loop of Henle. **C.** Distal convoluted tubular cell. **D.** Overview of entire nephron. **E.** Cortical collecting duct cells. **F.** Typical cell in the inner medullary collecting duct. The major membrane transporters, channels, and pumps are drawn with *arrows* indicating the direction of solute or water movement. For some events, the stoichiometry of transport is indicated by numerals preceding the solute. Targets for major diuretic agents are labeled. The actions of hormones are illustrated by *arrows with plus signs* for stimulatory effects and *lines with perpendicular ends* for inhibitory events. The *dashed line* indicates water impermeability of cell membranes in the thick ascending limb and distal convoluted tubule.



A

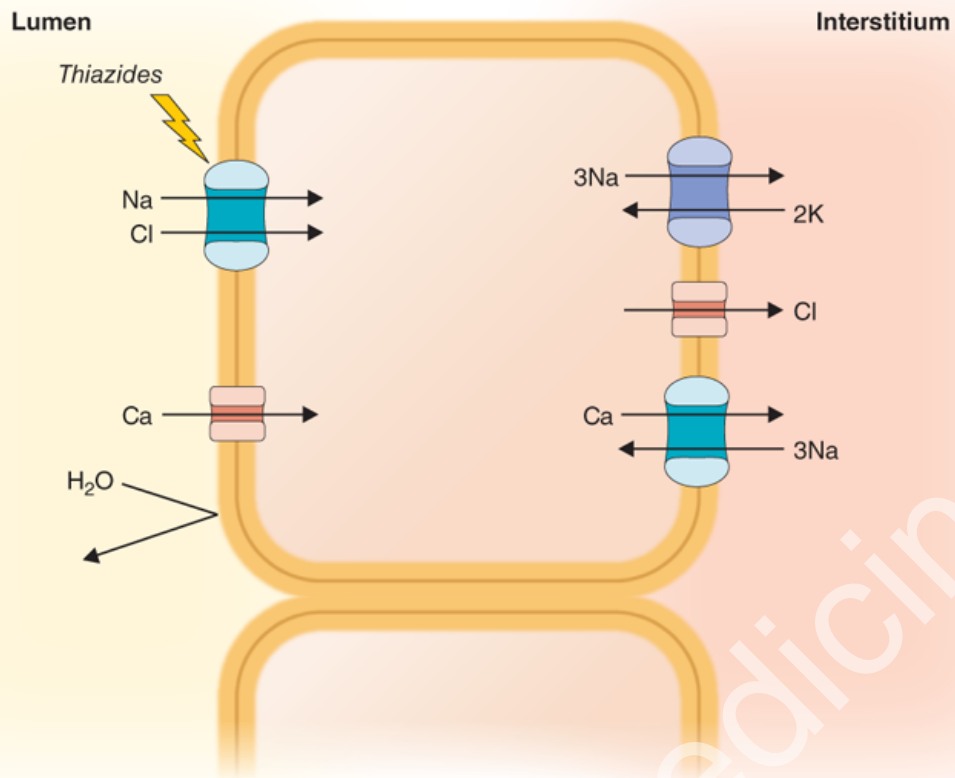
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

THICK ASCENDING LIMB

**B**

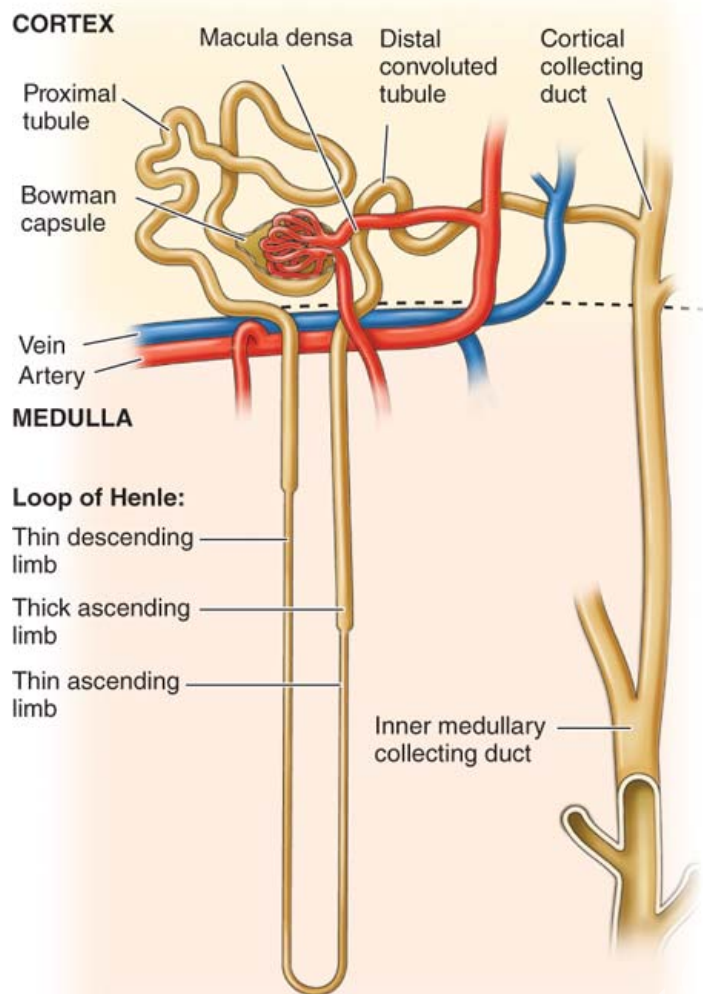
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
 Copyright © McGraw-Hill Education. All rights reserved.

DISTAL CONVOLUTED TUBULE



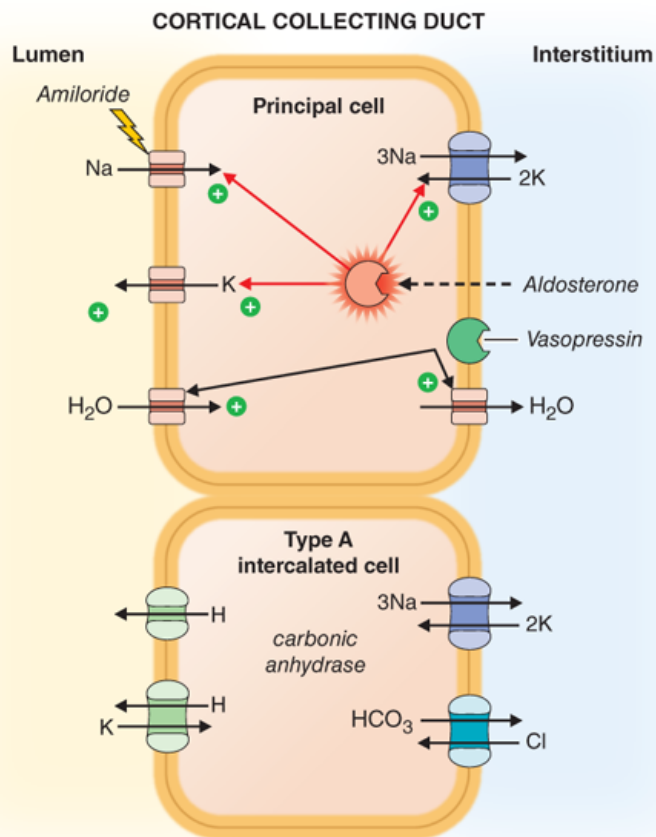
C

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

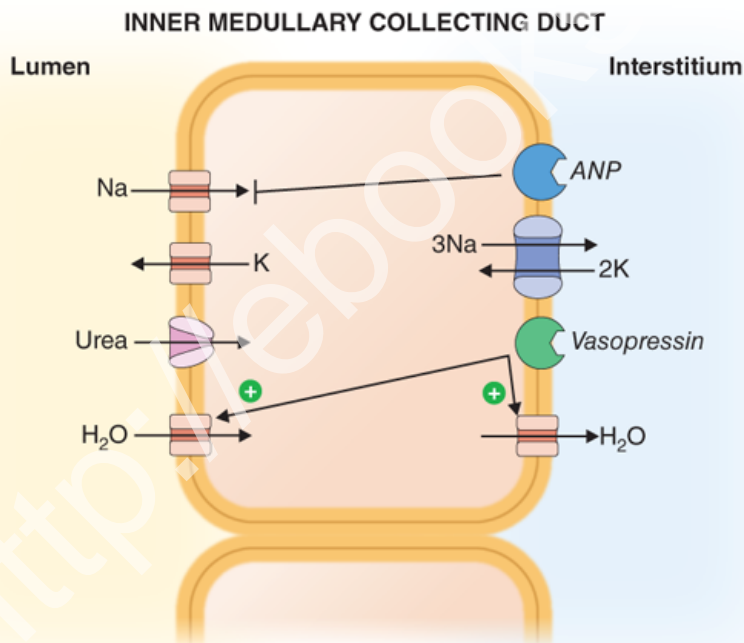


D

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

**E**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

**F**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

EPITHELIAL SOLUTE TRANSPORT

There are two types of epithelial transport. Movement of fluid and solutes sequentially across the apical and basolateral cell membranes (or vice versa) mediated by transporters, channels, or pumps is called *cellular transport*. By contrast, movement of fluid and solutes through the narrow passageway between adjacent cells is called *paracellular transport*. Paracellular transport occurs through tight junctions, indicating that they are not completely “tight” or occlusive. Indeed, some epithelial cell layers allow rather robust paracellular transport to occur (*leaky epithelia*), whereas other epithelia have more restrictive tight junctions (*tight epithelia*). In addition, because the ability of ions to flow through the paracellular pathway determines the electrical resistance across the epithelial monolayer, leaky and tight epithelia are also referred to as low- or high-resistance epithelia, respectively. The proximal tubule contains leaky epithelia, whereas distal nephron segments, such as the collecting duct, contain tight epithelia. Leaky epithelia are most well suited for bulk fluid reabsorption, whereas tight epithelia allow for more refined control and regulation of transport.

MEMBRANE TRANSPORT

Cell membranes are composed of hydrophobic lipids that repel water and aqueous solutes. The movement of solutes and water across cell membranes is made possible by discrete classes of integral membrane proteins, including channels, pumps, and transporters. These different mechanisms mediate specific types of transport activities, including *active transport* (pumps), *passive transport* (channels), *facilitated diffusion* (transporters), and *secondary active transport* (cotransporters). Active transport requires metabolic energy generated by the hydrolysis of ATP. Active transport pumps are ion-translocating ATPases, including the ubiquitous Na^+/K^+ -ATPase, the H^+ -ATPases, and Ca^{2+} -ATPases. Active transport creates asymmetric ion concentrations across a cell membrane and can move ions against a chemical gradient. The potential energy stored in a concentration gradient of an ion such as Na^+ can be used to drive transport through other mechanisms (secondary active transport). Pumps are often *electrogenic*, meaning they can create an asymmetric distribution of electrostatic charges across the membrane and establish a voltage or membrane potential. The movement of solutes through a membrane protein by simple diffusion is called passive transport. This activity is mediated by channels created by selectively permeable membrane proteins, and it allows solute or water to move across a membrane driven by favorable *concentration gradients* or *electrochemical potential*. Facilitated diffusion is a specialized type of passive transport mediated by simple transporters called *carriers* or *uniporters*. For example, hexose transporters such as GLUT2 mediate glucose transport by tubular cells. These transporters are driven by the concentration gradient for glucose that is highest in extracellular fluids and lowest in the cytoplasm due to rapid metabolism. Many other transporters operate by translocating two or more ions/solutes in concert either in the same direction (*symporters* or *cotransporters*) or in opposite directions (*antiporters* or *exchangers*) across the cell membrane. The movement of two or more ions/solutes may produce no net change in the balance of electrostatic charges across the membrane (*electroneutral*), or a transport event may alter the balance of charges (*electrogenic*). Several inherited disorders of renal tubular solute and water transport occur as a consequence of mutations in genes encoding a variety of channels, transporter proteins, and their regulators ([Table 303-1](#)).

TABLE 303-1

Inherited Disorders Affecting Renal Tubular Ion and Solute Transport

DISEASE OR SYNDROME	GENE	OMIM ^a
Disorders Involving the Proximal Tubule		
Proximal renal tubular acidosis	Sodium bicarbonate cotransporter (<i>SLC4A4</i> , 4q21)	604278
Fanconi-Bickel syndrome	Glucose transporter, GLUT2 (<i>SLC2A2</i> , 3q26.2)	227810
Isolated renal glycosuria	Sodium glucose cotransporter (<i>SLC5A2</i> , 16p11.2)	233100
Cystinuria		
Type I	Cystine, dibasic and neutral amino acid transporter (<i>SLC3A1</i> , 2p16.3)	220100
Non-type I	Amino acid transporter, light subunit (<i>SLC7A9</i> , 19q13.1)	600918
Lysinuric protein intolerance	Amino acid transporter (<i>SLC7A7</i> , 4q11.2)	222700
Hartnup disorder	Neutral amino acid transporter (<i>SLC6A19</i> , 5p15.33)	34500
Hereditary hypophosphatemic rickets with hypercalcemia	Sodium phosphate cotransporter (<i>SLC34A3</i> , 9q34)	241530
Renal hypouricemia		
Type 1	Urate-anion exchanger (<i>SLC22A12</i> , 11q13)	220150
Type 2	Urate transporter, GLUT9 (<i>SLC2A9</i> , 4p16.1)	612076
Dent disease	Chloride channel, CIC-5 (<i>CLCN5</i> , Xp11.22)	300009
X-linked recessive nephrolithiasis with renal failure	Chloride channel, CIC-5 (<i>CLCN5</i> , Xp11.22)	310468
X-linked recessive hypophosphatemic rickets	Chloride channel, CIC-5 (<i>CLCN5</i> , Xp11.22)	307800
Disorders Involving the Loop of Henle		
Bartter's syndrome		
Type 1	Sodium, potassium chloride cotransporter (<i>SLC12A1</i> , 15q21.1)	241200
Type 2	Potassium channel, ROMK (<i>KCNJ1</i> , 11q24)	601678
Type 3	Chloride channel, CIC-Kb (<i>CLCNKB</i> , 1p36)	602023
with sensorineural deafness	Chloride channel accessory subunit, Barttin (<i>BSND</i> , 1p31)	602522
Autosomal dominant hypocalcemia with Bartter-like syndrome	Calcium-sensing receptor (<i>CASR</i> , 3q13.33)	601199
Familial hypocalciuric hypercalcemia	Calcium-sensing receptor (<i>CASR</i> , 3q13.33)	145980
Primary hypomagnesemia	Claudin-16 or paracellin-1 (<i>CLDN16</i> or <i>PCLN1</i> , 3q27)	248250

DISEASE OR SYNDROME	GENE	OMIM ^a
Isolated renal magnesium loss	Sodium potassium ATPase, γ_1 -subunit (<i>ATP1G1</i> , 11q23)	154020
Disorders Involving the Distal Tubule and Collecting Duct		
Gitelman syndrome	Sodium chloride cotransporter (<i>SLC12A3</i> , 16q13)	263800
Primary hypomagnesemia with secondary hypocalcemia	Melastatin-related transient receptor potential cation channel 6 (<i>TRPM6</i> , 9q22)	602014
Pseudoaldosteronism (Liddle's syndrome)	Epithelial sodium channel β and γ subunits (<i>SCNN1B</i> , <i>SCNN1G</i> , 16p12.1)	177200
Recessive pseudohypoaldosteronism type 1	Epithelial sodium channel, α , β , and γ subunits (<i>SCNN1A</i> , 12p13; <i>SCNN1B</i> , <i>SCNN1G</i> , 16pp12.1)	264350
Pseudohypoaldosteronism type 2 (Gordon's hyperkalemia-hypertension syndrome)	Kinases WNK-1, WNK-4 (<i>WNK1</i> , 12p13; <i>WNK4</i> , 17q21.31)	145260
X-linked nephrogenic diabetes insipidus	Vasopressin V2 receptor (<i>AVPR2</i> , Xq28)	304800
Nephrogenic diabetes insipidus (autosomal)	Water channel, aquaporin-2 (<i>AQP2</i> , 12q13)	125800
Distal renal tubular acidosis		
autosomal dominant	Anion exchanger-1 (<i>SLC4A1</i> , 17q21.31)	179800
autosomal recessive	Anion exchanger-1 (<i>SLC4A1</i> , 17q21.31)	602722
with neural deafness	Proton ATPase, β_1 subunit (<i>ATP6V1B1</i> , 2p13.3)	192132
with normal hearing	Proton ATPase, 116-kD subunit (<i>ATP6V0A4</i> , 7q34)	602722

^aOnline Mendelian Inheritance in Man database (<http://www.ncbi.nlm.nih.gov.ezproxy.uaeu.ac.ae/Omim>).

SEGMENTAL NEPHRON FUNCTIONS

Each anatomic segment of the nephron has unique characteristics and specialized functions enabling selective transport of solutes and water (Fig. 303-3). Through sequential events of reabsorption and secretion along the nephron, tubular fluid is progressively conditioned into urine. Knowledge of the major tubular mechanisms responsible for solute and water transport is critical for understanding hormonal regulation of kidney function and the pharmacologic manipulation of renal excretion.

PROXIMAL TUBULE

The proximal tubule is responsible for reabsorbing ~60% of filtered NaCl and water, as well as ~90% of filtered bicarbonate and most critical nutrients such as glucose and amino acids. The proximal tubule uses both cellular and paracellular transport mechanisms. The apical membrane of proximal tubular cells has an expanded surface area available for reabsorptive work created by a dense array of microvilli called the *brush border*, and leaky tight junctions enable high-capacity fluid reabsorption.

Solute and water pass through these tight junctions to enter the lateral intercellular space where absorption by the peritubular capillaries occurs. Bulk fluid reabsorption by the proximal tubule is driven by high oncotic pressure and low hydrostatic pressure within the peritubular

capillaries. Cellular transport of most solutes by the proximal tubule is coupled to the Na^+ concentration gradient established by the activity of a basolateral Na^+/K^+ -ATPase (Fig. 303-3A). This active transport mechanism maintains a steep Na^+ gradient by keeping intracellular Na^+ concentrations low. Solute reabsorption from the tubular lumen is coupled to the Na^+ gradient by Na^+ -dependent transporters such as Na^+ -glucose and Na^+ -phosphate cotransporters present in apical membranes. In addition to the paracellular route, water reabsorption also occurs through the cellular pathway enabled by constitutively active water channels (aquaporin-1) present on both apical and basolateral membranes.

Proximal tubular cells reclaim nearly all filtered bicarbonate by a mechanism dependent on carbonic anhydrases. Filtered bicarbonate is first titrated by protons delivered to the lumen mainly by Na^+/H^+ exchange. The resulting carbonic acid (H_2CO_3) is metabolized by brush border carbonic anhydrase to water and carbon dioxide. Dissolved carbon dioxide then diffuses into the cell, where it is enzymatically hydrated by cytoplasmic carbonic anhydrase to re-form carbonic acid. Finally, intracellular carbonic acid dissociates into free protons and bicarbonate anions, and bicarbonate exits the cell through a basolateral $\text{Na}^+/\text{HCO}_3^-$ cotransporter. This process is saturable, which can result in urinary bicarbonate excretion when plasma levels exceed the physiologically normal range (24–26 meq/L). Carbonic anhydrase inhibitors such as acetazolamide, a class of weak diuretic agents, block proximal tubule bicarbonate reabsorption and are useful for alkalinizing the urine.

The proximal tubule contributes to acid secretion by two mechanisms involving the titration of the urinary buffers ammonia (NH_3) and phosphate. Renal NH_3 is produced by glutamine metabolism in the proximal tubule. Subsequent diffusion of NH_3 out of the proximal tubular cell enables trapping of H^+ , which is secreted by Na^+/H^+ exchange, in the lumen as ammonium ion (NH_4^+). Cellular K^+ levels inversely modulate proximal tubular ammoniogenesis, and in the setting of high serum K^+ from hypoaldosteronism, reduced ammoniogenesis promotes type IV renal tubular acidosis. Filtered hydrogen phosphate ion (HPO_4^{2-}) is also titrated in the proximal tubule by secreted H^+ to form H_2PO_4^- , and this reaction constitutes a major component of the urinary buffer referred to as titratable acid. Most filtered phosphate ion is reabsorbed by the proximal tubule through a sodium-coupled cotransport process that is regulated by parathyroid hormone (PTH).

Chloride is poorly reabsorbed throughout the first segment of the proximal tubule, and a rise in Cl^- concentration counterbalances the removal of bicarbonate anion from tubular fluid. In later proximal tubular segments, cellular Cl^- reabsorption is initiated by apical exchange of cellular formate for higher luminal concentrations of Cl^- . Once in the lumen, formate anions are titrated by H^+ (provided by Na^+/H^+ exchange) to generate neutral formic acid, which can diffuse passively across the apical membrane back into the cell where it dissociates a proton and is recycled. Basolateral Cl^- exit is mediated by a K^+/Cl^- cotransporter.

Reabsorption of glucose is nearly complete by the end of the proximal tubule. Cellular transport of glucose is mediated by apical Na^+ -glucose cotransport coupled with basolateral, facilitated diffusion by a glucose transporter. This process is also saturable, leading to glycosuria when plasma levels exceed 180–200 mg/dL, as seen in untreated diabetes mellitus.

The proximal tubule possesses specific transporters capable of secreting a variety of organic acids (carboxylate anions) and bases (mostly primary amine cations). Organic anions transported by these systems include urate, dicarboxylic acid anions (succinate), ketoacid anions, and several protein-bound drugs not filtered at the glomerulus (penicillins, cephalosporins, and salicylates). Probenecid inhibits renal organic anion secretion and can be clinically useful for raising plasma concentrations of certain drugs like penicillin and oseltamivir. Organic cations secreted by the proximal tubule include various biogenic amine neurotransmitters (dopamine, acetylcholine, epinephrine, norepinephrine, and histamine) and creatinine. The ATP-dependent transporter P-glycoprotein is highly expressed in brush border membranes and secretes several medically important drugs, including cyclosporine, digoxin, tacrolimus, and various cancer chemotherapeutic agents. Certain drugs like cimetidine and trimethoprim compete with endogenous compounds for transport by the organic cation pathways. Although these drugs elevate serum creatinine levels, there is no actual change in GFR in this setting.

Calcium and phosphorus homeostasis depends upon normal functioning of the proximal tubule. Approximately 60–70% of filtered calcium and ~85% of filtered phosphorus (in the form of inorganic phosphate) are reabsorbed by the proximal tubule. Whereas calcium reabsorption is mostly by passive diffusion through the paracellular route, phosphate reabsorption is mediated by sodium-coupled cotransport. In addition to direct reabsorption, the proximal tubule contributes to systemic mineral balance by participating in specific endocrine pathways. Circulating 25-hydroxy vitamin D (calcidiol) is bioactivated by proximal tubular 1α -hydroxylase to produce 1,25-di-hydroxy vitamin D (calcitriol), the most active form of the hormone, that acts on the small intestine to promote calcium absorption. Phosphate balance is affected by circulating fibroblast growth factor 23 (FGF23), a bone-derived hormone that interacts with its receptor (FGFR1) and co-receptor (Klotho) on proximal tubular cells to suppress sodium-phosphate cotransport and promote renal phosphate excretion. PTH stimulates proximal tubular 1α -

hydroxylation of vitamin D while it suppresses sodium-phosphate cotransport. Derangements in PTH and FGF23 account for abnormal calcium and phosphate balance in chronic kidney disease.

The proximal tubule, through distinct classes of Na^+ -dependent and Na^+ -independent transport systems, reabsorbs amino acids efficiently. These transporters are specific for different groups of amino acids. For example, cystine, lysine, arginine, and ornithine are transported by a system comprising two proteins encoded by the *SLC3A1* and *SLC7A9* genes. Mutations in either *SLC3A1* or *SLC7A9* impair reabsorption of these amino acids and cause the disease cystinuria. Peptide hormones, such as insulin and growth hormone, β_2 -microglobulin, albumin, and other small proteins, are taken up by the proximal tubule through a process of absorptive endocytosis and are degraded in acidified endocytic lysosomes. Acidification of these vesicles depends on a vacuolar H^+ -ATPase and Cl^- channel. Impaired acidification of endocytic vesicles because of mutations in a Cl^- channel gene (*CLCN5*) causes low-molecular-weight proteinuria in Dent disease.

LOOP OF HENLE

The loop of Henle consists of three major segments: descending thin limb, ascending thin limb, and ascending thick limb. These divisions are based on cellular morphology and anatomic location, but also correlate with specialization of function. Approximately 15–25% of filtered NaCl is reabsorbed in the loop of Henle, mainly by the thick ascending limb. The loop of Henle has an important role in urinary concentration by contributing to the generation of a hypertonic medullary interstitium in a process called *countercurrent multiplication*. The loop of Henle is the site of action for the most potent class of diuretic agents (loop diuretics) and also contributes to reabsorption of calcium and magnesium ions.

The descending thin limb is highly water permeable owing to dense expression of constitutively active aquaporin-1 water channels. By contrast, water permeability is negligible in the ascending limb. In the thick ascending limb, there is a high level of secondary active NaCl transport enabled by the $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ cotransporter on the apical membrane in series with basolateral Cl^- channels and Na^+/K^+ -ATPase (Fig. 303-3B). The $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ cotransporter is the primary target for loop diuretics. Tubular fluid K^+ is the limiting substrate for this cotransporter (tubular concentration of K^+ is similar to plasma, about 4 meq/L), but transporter activity is maintained by K^+ recycling through an apical potassium channel. The cotransporter also enables reabsorption of NH_4^+ in lieu of K^+ , and this leads to accumulation of both NH_4^+ and NH_3 in the medullary interstitium. An inherited disorder of the thick ascending limb, Bartter's syndrome, also results in a salt-wasting renal disease associated with hypokalemia and metabolic alkalosis. Loss-of-function mutations in one of five distinct genes encoding components of the $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ cotransporter (*NKCC2*), apical K^+ channel (*KCNJ1*), basolateral Cl^- channel (*CLCNKB*, *BSND*), or calcium-sensing receptor (*CASR*) can cause Bartter's syndrome.

Potassium recycling also contributes to a positive electrostatic charge in the lumen relative to the interstitium that promotes divalent cation (Mg^{2+} and Ca^{2+}) reabsorption through a paracellular pathway. A Ca^{2+} -sensing, G-protein-coupled receptor (CaSR) on basolateral membranes regulates NaCl reabsorption in the thick ascending limb through dual signaling mechanisms using either cyclic AMP or eicosanoids. This receptor enables a steep relationship between plasma Ca^{2+} levels and renal Ca^{2+} excretion. Loss-of-function mutations in CaSR cause familial hypercalcemic hypocalciuria because of a blunted response of the thick ascending limb to extracellular Ca^{2+} . Mutations in *CLDN16* encoding paracellin-1, a transmembrane protein located within the tight junction complex, leads to familial hypomagnesemia with hypercalciuria and nephrocalcinosis, suggesting that the ion conductance of the paracellular pathway in the thick limb is regulated.

The loop of Henle contributes to urine-concentrating ability by establishing a *hypertonic medullary interstitium* that promotes water reabsorption by the downstream inner medullary collecting duct. *Countercurrent multiplication* produces a hypertonic medullary interstitium using two countercurrent systems: the loop of Henle (opposing descending and ascending limbs) and the vasa recta (medullary peritubular capillaries enveloping the loop). The countercurrent flow in these two systems helps maintain the hypertonic environment of the inner medulla, but NaCl reabsorption by the thick ascending limb is the primary initiating event. Reabsorption of NaCl without water dilutes the tubular fluid and adds new osmoles to medullary interstitial fluid. Because the descending thin limb is highly water permeable, osmotic equilibrium occurs between the descending limb tubular fluid and the interstitial space, leading to progressive solute trapping in the inner medulla. Maximum medullary interstitial osmolality also requires partial recycling of urea from the collecting duct.

DISTAL CONVOLUTED TUBULE

The distal convoluted tubule reabsorbs ~5% of the filtered NaCl. This segment is composed of a tight epithelium with little water permeability. The major NaCl-transporting pathway uses an apical membrane, electroneutral thiazide-sensitive Na^+/Cl^- cotransporter in tandem with

basolateral Na^+/K^+ -ATPase and Cl^- channels (**Fig. 303-3C**). Apical Ca^{2+} -selective channels (TRPV5) and basolateral $\text{Na}^+/\text{Ca}^{2+}$ exchange mediate calcium reabsorption in the distal convoluted tubule. Ca^{2+} reabsorption is inversely related to Na^+ reabsorption and is stimulated by PTH. Blocking apical Na^+/Cl^- cotransport will reduce intracellular Na^+ , favoring increased basolateral $\text{Na}^+/\text{Ca}^{2+}$ exchange and passive apical Ca^{2+} entry. Loss-of-function mutations of *SLC12A3* encoding the apical Na^+/Cl^- cotransporter cause Gitelman syndrome, a salt-wasting disorder associated with hypokalemic alkalosis and hypocalciuria. Mutations in genes encoding WNK kinases, WNK-1 and WNK-4, cause pseudohypoaldosteronism type II (Gordon syndrome) characterized by familial hypertension with hyperkalemia. WNK kinases influence the activity of several tubular ion transporters. Mutations in this disorder lead to overactivity of the apical Na^+/Cl^- cotransporter in the distal convoluted tubule as the primary stimulus for increased salt reabsorption, extracellular volume expansion, and hypertension. Hyperkalemia may be caused by diminished activity of apical K^+ channels in the collecting duct, a primary route for K^+ secretion. Mutations in *TRPM6* encoding Mg^{2+} permeable ion channels also cause familial hypomagnesemia with hypocalcemia. A molecular complex of TRPM6 and TRPM7 proteins is critical for Mg^{2+} reabsorption in the distal convoluted tubule.

COLLECTING DUCT

The collecting duct modulates the final composition of urine. The two major divisions, the cortical collecting duct and inner medullary collecting duct, contribute to reabsorbing ~4–5% of filtered Na^+ and are important for hormonal regulation of salt and water balance. Cells in both segments of the collecting duct express vasopressin-regulated water channels (aquaporin-2 on the apical membrane, aquaporin-3 and -4 on the basolateral membrane). The antidiuretic hormone vasopressin binds to the V2 receptor on the basolateral membrane and triggers an intracellular signaling cascade through G-protein-mediated activation of adenylyl cyclase, resulting in an increase in the cellular levels of cyclic AMP. This signaling cascade stimulates the insertion of water channels into the apical membrane of collecting duct cells to promote increased water permeability. This increase in permeability enables water reabsorption and production of concentrated urine. In the absence of vasopressin, collecting duct cells are water impermeable, and urine remains dilute.

The cortical collecting duct contains *high-resistance epithelia* with two cell types. Principal cells are the main water, Na^+ -reabsorbing, and K^+ -secreting cells, and the site of action of aldosterone, K^+ -sparing diuretics, and mineralocorticoid receptor antagonists such as spironolactone and eplerenone. The other cells are type A and B intercalated cells. Type A intercalated cells mediate acid secretion and bicarbonate reabsorption also under the influence of aldosterone. Type B intercalated cells mediate bicarbonate secretion and acid reabsorption.

Virtually all transport is mediated through the cellular pathway for both principal cells and intercalated cells. In principal cells, passive apical Na^+ entry occurs through an amiloride-sensitive, epithelial Na^+ channel (ENaC) with basolateral exit mediated by the Na^+/K^+ -ATPase (**Fig. 303-3D**). This Na^+ reabsorptive process is tightly regulated by aldosterone and is physiologically activated by a variety of proteolytic enzymes that cleave extracellular domains of ENaC; plasmin in the tubular fluid of nephrotic patients, for example, activates ENaC, leading to sodium retention. Aldosterone enters the cell across the basolateral membrane, binds to a cytoplasmic mineralocorticoid receptor, and then translocates into the nucleus, where it modulates gene transcription, resulting in increased Na^+ reabsorption and K^+ secretion. Activating mutations in ENaC increase Na^+ reclamation and produce hypokalemia, hypertension, and metabolic alkalosis (Liddle's syndrome). The potassium-sparing diuretics amiloride and triamterene block ENaC, causing reduced Na^+ reabsorption.

Principal cells secrete K^+ through an apical membrane potassium channel. Several forces govern the secretion of K^+ . Most importantly, the high intracellular K^+ concentration generated by Na^+/K^+ -ATPase creates a favorable concentration gradient for K^+ secretion into tubular fluid. With reabsorption of Na^+ without an accompanying anion, the tubular lumen becomes negative relative to the cell interior, creating a favorable electrical gradient for secretion of potassium. When Na^+ reabsorption is blocked, the electrical component of the driving force for K^+ secretion is blunted, and this explains lack of excess urinary K^+ loss during treatment with potassium-sparing diuretics or mineralocorticoid receptor antagonists. K^+ secretion is also promoted by aldosterone actions that increase regional Na^+ transport, which favor more lumen electronegativity, and by increasing the number and activity of potassium channels. Fast tubular fluid flow rates that occur during volume expansion or diuretics acting “upstream” of the cortical collecting duct also increase K^+ secretion, as does the presence of relatively nonreabsorbable anions (including bicarbonate and semisynthetic penicillins) that contribute to the lumen-negative potential. Off-target effects of certain antibiotics, such as trimethoprim and **pentamidine**, block ENaCs and predispose to hyperkalemia, especially when renal K^+ handling is impaired for other reasons. Principal cells, as described below, also participate in water reabsorption by increased water permeability in response to vasopressin.

Intercalated cells do not participate in Na^+ reabsorption but, instead, mediate acid-base secretion. These cells perform two types of transport: active H^+ transport mediated by H^+ -ATPase (proton pump), and $\text{Cl}^-/\text{HCO}_3^-$ exchange. Intercalated cells arrange the two transport mechanisms on opposite membranes to enable either acid or base secretion. Type A intercalated cells have an apical proton pump that mediates acid secretion and a basolateral $\text{Cl}^-/\text{HCO}_3^-$ anion exchanger for bicarbonate reabsorption (Fig. 303-3E). Aldosterone increases the number of H^+ -ATPase pumps, sometimes contributing to the development of metabolic alkalosis. Secreted H^+ is buffered by NH_3 that has diffused into the collecting duct lumen from the surrounding interstitium. By contrast, type B intercalated cells have the $\text{Cl}^-/\text{HCO}_3^-$ exchanger on the apical membrane to mediate bicarbonate secretion while the proton pump resides on the basolateral membrane to enable acid reabsorption. Under conditions of acidemia, the kidney preferentially uses type A intercalated cells to secrete the excess H^+ and generate more HCO_3^- . The opposite is true in states of bicarbonate excess with alkalemia where the type B intercalated cells predominate. An extracellular protein called *hensin* mediates this adaptation.

Inner medullary collecting duct cells share many similarities with principal cells of the cortical collecting duct. They have apical Na^+ and K^+ channels that mediate Na^+ reabsorption and K^+ secretion, respectively (Fig. 303-3F). Sodium reabsorption by inner medullary collecting duct cells is also inhibited by the natriuretic peptides called *atrial natriuretic peptide* or *renal natriuretic peptide* (urodilatin); the same gene encodes both peptides but uses different posttranslational processing of a common prehormone to generate different proteins. Atrial natriuretic peptides are secreted by atrial myocytes in response to volume expansion, whereas urodilatin is secreted by renal tubular epithelia. Natriuretic peptides interact with either apical (urodilatin) or basolateral (atrial natriuretic peptides) receptors on inner medullary collecting duct cells to stimulate guanylyl cyclase and increase levels of cytoplasmic cGMP. This effect in turn reduces the activity of the apical Na^+ channel in these cells and attenuates net Na^+ reabsorption, producing natriuresis.

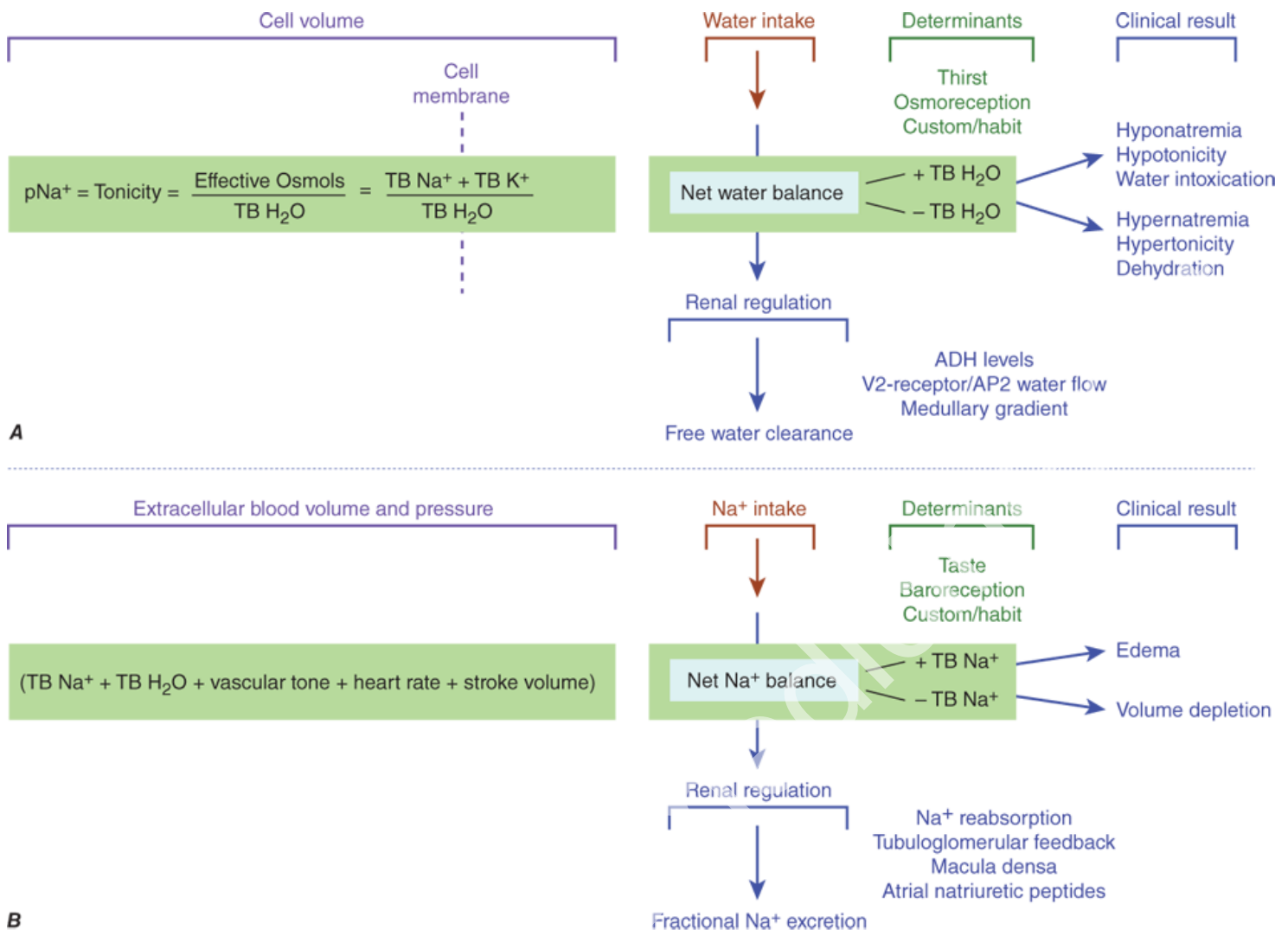
The inner medullary collecting duct transports *urea* out of the lumen, returning *urea* to the interstitium, where it contributes to the hypertonicity of the medullary interstitium. *Urea* is recycled by diffusing from the interstitium into the descending and ascending limbs of the loop of Henle.

HORMONAL REGULATION OF SODIUM AND WATER BALANCE

The balance of solute and water in the body is determined by the amounts ingested, distributed to various fluid compartments, and excreted by skin, bowel, and kidneys. *Tonicity*, the osmolar state determining the volume behavior of cells in a solution, is regulated by water balance (Fig. 303-4A), and *extracellular blood volume* is regulated by Na^+ balance (Fig. 303-4B). The kidney is a critical modulator of both physiologic processes.

FIGURE 303-4

Determinants of sodium and water balance. **A.** Plasma Na^+ concentration is a surrogate marker for plasma tonicity, the volume behavior of cells in a solution. Tonicity is determined by the number of effective osmoles in the body divided by the total body H_2O (TB H_2O), which translates simply into the total body Na (TB Na^+) and anions outside the cell separated from the total body K (TB K^+) inside the cell by the cell membrane. Net water balance is determined by the integrated functions of thirst, osmoreception, Na reabsorption, vasopressin release, and the strength of the medullary gradient in the kidney, keeping tonicity within a narrow range of osmolality around 280 mosmol/L. When water metabolism is disturbed and total body water increases, hyponatremia, hypotonicity, and water intoxication occur; when total body water decreases, hypernatremia, hypertonicity, and dehydration occur. **B.** Extracellular blood volume and pressure are an integrated function of total body Na^+ (TB Na^+), total body H_2O (TB H_2O), vascular tone, heart rate, and stroke volume that modulates volume and pressure in the vascular tree of the body. This extracellular blood volume is determined by net Na balance under the control of taste, baroreception, habit, Na^+ reabsorption, macula densa/tubuloglomerular feedback, and natriuretic peptides. When Na^+ metabolism is disturbed and total body Na^+ increases, edema occurs; when total body Na^+ is decreased, volume depletion occurs. ADH, antidiuretic hormone; AQP2, aquaporin-2.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

WATER BALANCE

Tonicity depends on the variable concentration of *effective osmoles* inside and outside the cell causing water to move in either direction across its membrane. Classic effective osmoles, like Na⁺, K⁺, and their anions, are solutes trapped on either side of a cell membrane, where they collectively partition and obligate water to move and find equilibrium in proportion to retained solute; Na⁺/K⁺-ATPase keeps most K⁺ inside cells and most Na⁺ outside. Normal tonicity (~280 mosmol/L) is rigorously defended by osmoregulatory mechanisms that control water balance to protect tissues from inadvertent *dehydration* (cell shrinkage) or *water intoxication* (cell swelling), both of which are deleterious to cell function (Fig. 303-4A).

The mechanisms that control osmoregulation are distinct from those governing extracellular volume, although there is some shared physiology in both processes. While cellular concentrations of K⁺ have a determinant role in any level of tonicity, the routine surrogate marker for assessing clinical tonicity is the concentration of serum Na⁺. Any reduction in total body water, which raises the Na⁺ concentration, triggers a brisk sense of thirst and conservation of water by decreasing renal water excretion mediated by release of vasopressin from the posterior pituitary. Conversely, a decrease in plasma Na⁺ concentration triggers an increase in renal water excretion by suppressing the secretion of vasopressin. Whereas all cells expressing mechanosensitive TRPV1, 2, or 4 channels, among potentially other sensors, respond to changes in tonicity by altering their volume and Ca²⁺ concentration, only TRPV⁺ neuronal cells connected to the organum vasculosum of the lamina terminalis are *osmoreceptive*. Only these cells, because of their neural connectivity and adjacency to a minimal blood-brain barrier, modulate the downstream release of vasopressin by the posterior lobe of the pituitary gland. Secretion is stimulated primarily by changing tonicity and secondarily by other nonosmotic signals such as variable blood volume, stress, pain, nausea, and some drugs. The release of vasopressin by the posterior pituitary increases linearly as plasma tonicity rises above normal, although this varies, depending on the perception of

extracellular volume (one form of cross-talk between mechanisms that adjudicate blood volume and osmoregulation). Changing the intake or excretion of water provides a means for adjusting plasma tonicity; thus, osmoregulation governs water balance.

The kidneys play a vital role in maintaining water balance through the regulation of renal water excretion. The ability to concentrate urine to an osmolality exceeding that of plasma enables water conservation, whereas the ability to produce urine more dilute than plasma promotes excretion of excess water. For water to enter or exit a cell, the cell membrane must express aquaporins. In the kidney, aquaporin-1 is constitutively active in all water-permeable segments (e.g., proximal tubule, descending thin limb of the loop of Henle), whereas aquaporin-2, -3, and -4 in the collecting duct promote vasopressin-regulated water permeability. Net water reabsorption is ultimately driven by the osmotic gradient between dilute tubular fluid and a hypertonic medullary interstitium.

SODIUM BALANCE

The perception of *extracellular blood volume* is determined, in part, by the integration of arterial tone, cardiac stroke volume, heart rate, and the water and solute content of extracellular fluid. Na^+ and accompanying anions are the most abundant extracellular effective osmoles and together support a blood volume around which pressure is generated. Under normal conditions, this volume is regulated by sodium balance (Fig. 303-4B), and the balance between daily Na^+ intake and excretion is under the influence of *baroreceptors* in regional blood vessels and vascular hormone sensors modulated by atrial natriuretic peptides, the renin-angiotensin-aldosterone system, Ca^{2+} signaling, adenosine, vasopressin, and the neural adrenergic axis. If Na^+ intake exceeds Na^+ excretion (positive Na^+ balance), then an increase in blood volume will trigger a proportional increase in urinary Na^+ excretion. Conversely, when Na^+ intake is less than urinary excretion (negative Na^+ balance), blood volume will decrease and trigger enhanced renal Na^+ reabsorption, leading to decreased urinary Na^+ excretion.

The renin-angiotensin-aldosterone system is the best-understood hormonal system modulating renal Na^+ excretion. Renin is synthesized and secreted by granular cells in the wall of the afferent arteriole. Its secretion is controlled by several factors, including β_1 -adrenergic stimulation to the afferent arteriole, input from the macula densa, and prostaglandins. Renin and ACE activity eventually produce angiotensin II that directly and indirectly promotes renal Na^+ and water reabsorption. Stimulation of proximal tubular Na^+/H^+ exchange by angiotensin II directly increases Na^+ reabsorption. Angiotensin II also promotes Na^+ reabsorption along the collecting duct by stimulating aldosterone secretion by the adrenal cortex. Constriction of the efferent glomerular arteriole by angiotensin II indirectly increases the filtration fraction and raises peritubular capillary oncotic pressure to promote tubular Na^+ reabsorption. Finally, angiotensin II inhibits renin secretion through a negative feedback loop. Alternative metabolism of angiotensin by ACE2 generates the vasodilatory peptide angiotensin 1-7 that acts through Mas receptors to counterbalance several actions of angiotensin II on blood pressure and renal function (Fig. 303-2C).

Aldosterone is synthesized and secreted by granulosa cells in the adrenal cortex. It binds to cytoplasmic mineralocorticoid receptors in the collecting duct principal cells that increase activity of ENaC, apical membrane K^+ channel, and basolateral Na^+/K^+ -ATPase. These effects are mediated in part by aldosterone-stimulated transcription of the gene encoding serum/glucocorticoid-induced kinase 1 (SGK1). The activity of ENaC is increased by SGK1-mediated phosphorylation of Nedd4-2, a protein that promotes recycling of the Na^+ channel from the plasma membrane. Phosphorylated Nedd4-2 has impaired interactions with ENaC, leading to increased channel density at the plasma membrane and increased capacity for Na^+ reabsorption by the collecting duct.

Chronic exposure to aldosterone causes a decrease in urinary Na^+ excretion lasting only a few days, after which Na^+ excretion returns to previous levels. This phenomenon, called *aldosterone escape*, is explained by decreased proximal tubular Na^+ reabsorption following blood volume expansion. Excess Na^+ that is not reabsorbed by the proximal tubule overwhelms the reabsorptive capacity of more distal nephron segments. This escape may be facilitated by atrial natriuretic peptides that lose their effectiveness in the clinical settings of heart failure, nephrotic syndrome, and cirrhosis, leading to severe Na^+ retention and volume overload.

FURTHER READING

Bhave G, Neilson EG: Body fluid dynamics: Back to the future. J Am Soc Nephrol 22:2166, 2011. [PubMed: 22034644]

Carlstrom M et al: Renal autoregulation in health and disease. Physiol Rev 95:405, 2015. [PubMed: 25834230]

Combes AN et al: Cell-cell interactions driving kidney morphogenesis. Curr Top Dev Biol 112:467, 2015. [PubMed: 25733149]

Curthoys NP, Moe OW: Proximal tubule function and response to acidosis. Clin J Am Soc Nephrol 9:1627, 2014. [[PubMed: 23908456](#)]

Hadchouel J et al: Regulation of renal electrolyte transport by WNK and SPAK-OSR1 kinases. Annu Rev Physiol 78:367, 2016. [[PubMed: 26863326](#)]

Jung HJ, Kwon TH: Molecular mechanisms regulating aquaporin-2 in kidney collecting duct. Am J Physiol Renal Physiol 331:F1318, 2016.

Mount DB: Thick ascending limb of the loop of Henle. Clin J Am Soc Nephrol 9:1974, 2014. [[PubMed: 25318757](#)]

Nigam SK et al: Handling of drugs, metabolites, and uremic toxins by kidney proximal tubule drug transporters. Clin J Am Soc Nephrol 10:2039, 2015. [[PubMed: 26490509](#)]

Palmer LG, Schnermann J: Integrated control of Na transport along the nephron. Clin J Am Soc Nephrol 10:676, 2015. [[PubMed: 25098598](#)]

Pearce D et al: Collecting duct principal cell transport processes and their regulation. Clin J Am Soc Nephrol 10:135, 2015. [[PubMed: 24875192](#)]

Terker AS, Ellison DH: Renal mineralocorticoid receptor and electrolyte homeostasis. Am J Physiol Regul Integr Comp Physiol 309:R1068, 2015. [[PubMed: 26136532](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 332: Acute Viral Hepatitis

Jules L. Dienstag

INTRODUCTION

Acute viral hepatitis is a systemic infection affecting the liver predominantly. Almost all cases of acute viral hepatitis are caused by one of five viral agents: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the HBV-associated delta agent or hepatitis D virus (HDV), and hepatitis E virus (HEV). All these human hepatitis viruses are RNA viruses, except for hepatitis B, which is a DNA virus but replicates like a retrovirus. Although these agents can be distinguished by their molecular and antigenic properties, all types of viral hepatitis produce clinically similar illnesses. These range from asymptomatic and inapparent to fulminant and fatal acute infections common to all types, on the one hand, and from subclinical persistent infections to rapidly progressive chronic liver disease with cirrhosis and even hepatocellular carcinoma, common to the bloodborne types (HBV, HCV, and HDV), on the other.

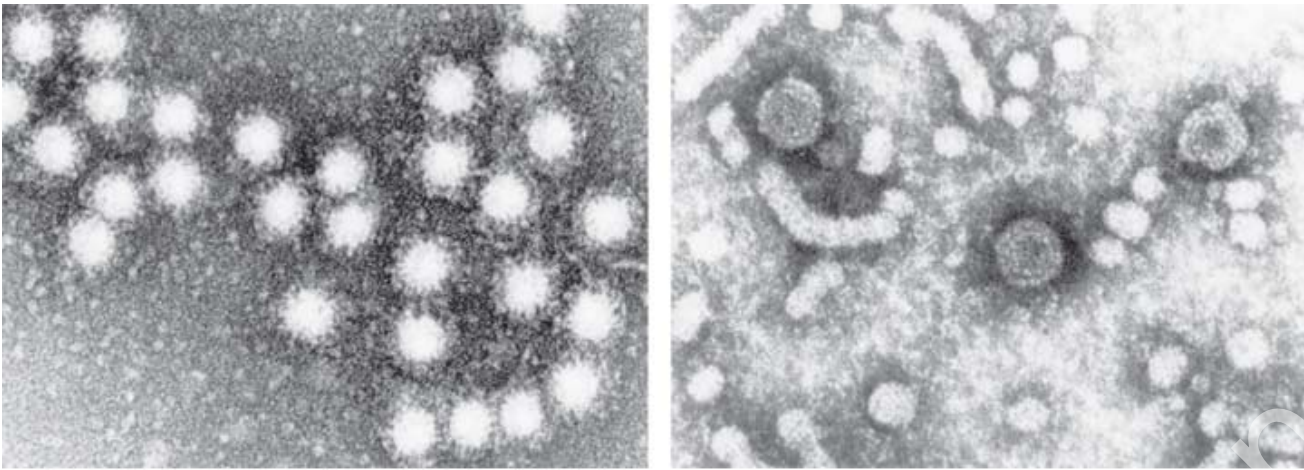
VIROLOGY AND ETIOLOGY

Hepatitis A

HAV is a nonenveloped 27-nm, heat-, acid-, and ether-resistant RNA virus in the *Hepatovirus* genus of the picornavirus family (**Fig. 332-1**). Its virion contains four capsid polypeptides, designated VP1–VP4, which are cleaved posttranslationally from the polyprotein product of a 7500-nucleotide genome. Inactivation of viral activity can be achieved by boiling for 1 min, by contact with formaldehyde and chlorine, or by ultraviolet irradiation. Despite nucleotide sequence variation of up to 20% among isolates of HAV, and despite the recognition of four genotypes affecting humans, all strains of this virus are immunologically indistinguishable and belong to one serotype. Human HAV can infect and cause hepatitis in chimpanzees, tamarins (marmosets), and several monkey species. Recently, a hepatotropic *Hepatovirus* related to, and likely to have shared common evolutionary ancestry with, human HAV has been identified in several species of harbor seals, albeit without histologic evidence for liver injury or inflammation. Hepatitis A has an incubation period of ~4 weeks. Its replication is limited to the liver, but the virus is present in the liver, bile, stools, and blood during the late incubation period and acute preicteric/presymptomatic phase of illness. Despite slightly longer persistence of virus in the liver, fecal shedding, viremia, and infectivity diminish rapidly once jaundice becomes apparent. HAV can be cultivated reproducibly in vitro.

FIGURE 332-1

Electron micrographs of hepatitis A virus particles and serum from a patient with hepatitis B. *Left*: 27-nm hepatitis A virus particles purified from stool of a patient with acute hepatitis A and aggregated by antibody to hepatitis A virus. ***Right*:** Concentrated serum from a patient with hepatitis B, demonstrating the 42-nm virions, tubular forms, and spherical 22-nm particles of hepatitis B surface antigen. 132,000×. (Hepatitis D resembles 42-nm virions of hepatitis B but is smaller, 35–37 nm; hepatitis E resembles hepatitis A virus but is slightly larger, 32–34 nm; hepatitis C has been visualized as a 55-nm particle.)

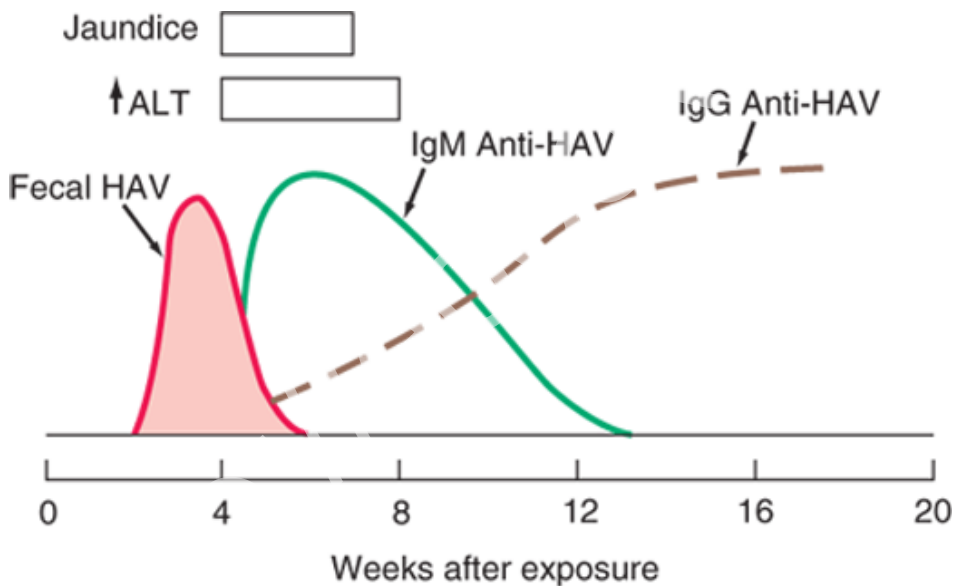


Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Antibodies to HAV (anti-HAV) can be detected during acute illness when serum aminotransferase activity is elevated and fecal HAV shedding is still occurring. This early antibody response is predominantly of the IgM class and persists for several (~3) months, rarely for 6–12 months. During convalescence, however, anti-HAV of the IgG class becomes the predominant antibody (Fig. 332-2). Therefore, the diagnosis of hepatitis A is made during acute illness by demonstrating anti-HAV of the IgM class. After acute illness, anti-HAV of the IgG class remains detectable indefinitely, and patients with serum anti-HAV are immune to reinfection. Neutralizing antibody activity parallels the appearance of anti-HAV, and the IgG anti-HAV present in [immune globulin](#) accounts for the protection it affords against HAV infection.

FIGURE 332-2

Scheme of typical clinical and laboratory features of hepatitis A virus (HAV). ALT, alanine aminotransferase.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

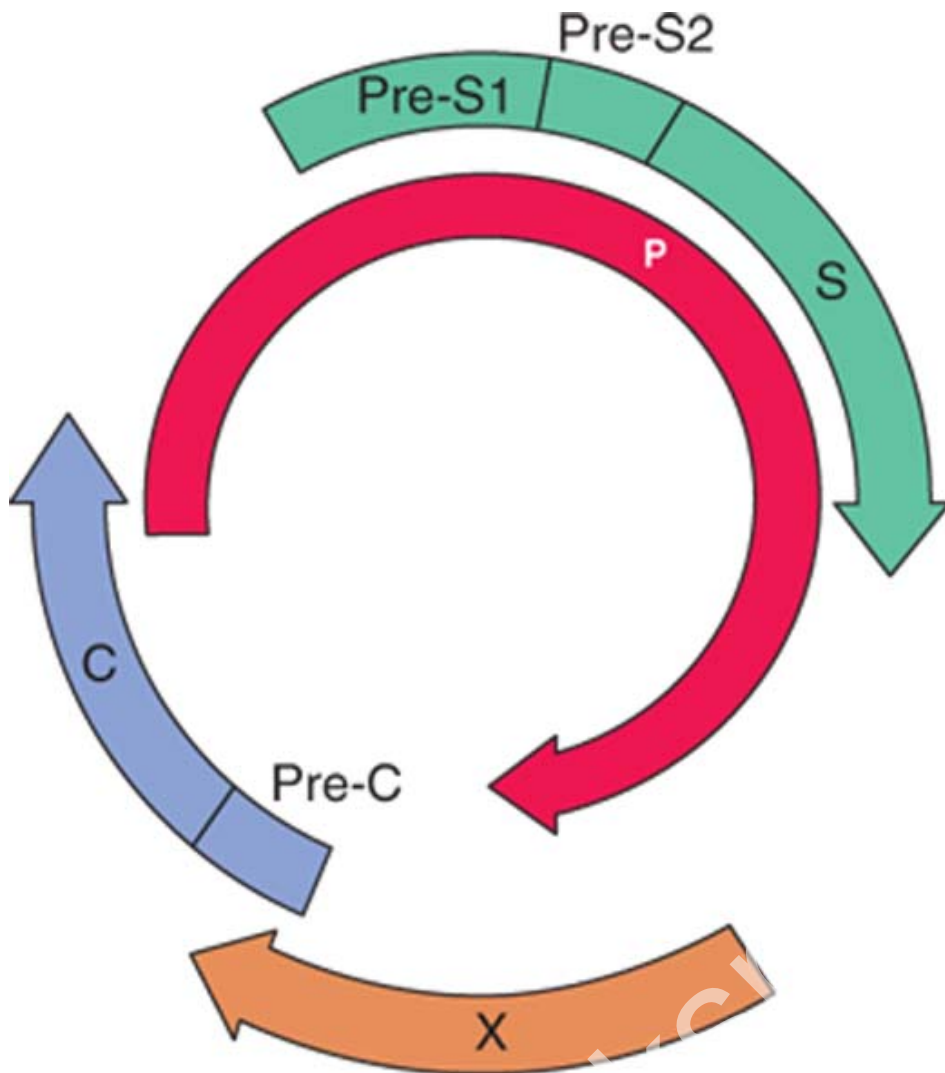
Hepatitis B

HBV is a DNA virus with a remarkably compact genomic structure; despite its small, circular, 3200-bp size, HBV DNA codes for four sets of viral products with a complex, multiparticulate structure. HBV achieves its genomic economy by relying on an efficient strategy of encoding proteins from four overlapping genes: S, C, P, and X (Fig. 332-3), as detailed

below. Once thought to be unique among viruses, HBV is now recognized as one of a family of animal viruses, hepadnaviruses (hepatotropic DNA viruses), and is classified as hepadnavirus type 1. Similar viruses infect certain species of woodchucks, ground and tree squirrels, and Pekin ducks, to mention the most carefully characterized; genetic evidence of ancient HBV-like virus forbears has been found in fossils of ancient birds, and a HBV-like virus has been identified in contemporary fish. Like HBV, all have the same distinctive three morphologic forms, have counterparts to the envelope and nucleocapsid virus antigens of HBV, replicate in the liver but exist in extrahepatic sites, contain their own endogenous DNA polymerase, have partially double-strand and partially single-strand genomes, are associated with acute and chronic hepatitis and hepatocellular carcinoma, and rely on a replicative strategy unique among DNA viruses but typical of retroviruses. Entry of HBV into hepatocytes is mediated by binding to the sodium taurocholate cotransporting polypeptide receptor. Instead of DNA replication directly from a DNA template, hepadnaviruses rely on reverse transcription (effected by the DNA polymerase) of minus-strand DNA from a “pregenomic” RNA intermediate. Then, plus-strand DNA is transcribed from the minus-strand DNA template by the DNA-dependent DNA polymerase and converted in the hepatocyte nucleus to a covalently closed circular DNA, which serves as a template for messenger RNA and pregenomic RNA. Viral proteins are translated by the messenger RNA, and the proteins and genome are packaged into virions and secreted from the hepatocyte. Although HBV is difficult to cultivate *in vitro* in the conventional sense from clinical material, several cell lines have been transfected with HBV DNA. Such transfected cells support *in vitro* replication of the intact virus and its component proteins.

FIGURE 332-3

Compact genomic structure of hepatitis B virus (HBV). This structure, with overlapping genes, permits HBV to code for multiple proteins. The S gene codes for the “major” envelope protein, HBsAg. Pre-S1 and pre-S2, upstream of S, combine with S to code for two larger proteins, “middle” protein, the product of pre-S2 + S, and “large” protein, the product of pre-S1 + pre-S2 + S. The largest gene, P, codes for DNA polymerase. The C gene codes for two nucleocapsid proteins, HBeAg, a soluble, secreted protein (initiation from the pre-C region of the gene), and HBcAg, the intracellular core protein (initiation after pre-C). The X gene codes for HBxAg, which can transactivate the transcription of cellular and viral genes; its clinical relevance is not known, but it may contribute to carcinogenesis by binding to p53.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

VIRAL PROTEINS AND PARTICLES

Of the three particulate forms of HBV (**Table 332-1**), the most numerous are the 22-nm particles, which appear as spherical or long filamentous forms; these are antigenically indistinguishable from the outer surface or envelope protein of HBV and are thought to represent excess viral envelope protein. Outnumbered in serum by a factor of 100 or 1000 to 1 compared with the spheres and tubules are large, 42-nm, double-shelled spherical particles, which represent the intact hepatitis B virion (**Fig. 332-1**). The envelope protein expressed on the outer surface of the virion and on the smaller spherical and tubular structures is referred to as *hepatitis B surface antigen* (HBsAg). The concentration of HBsAg and virus particles in the blood may reach 500 µg/mL and 10 trillion particles per milliliter, respectively. The envelope protein, HBsAg, is the product of the S gene of HBV.

TABLE 332-1

Nomenclature and Features of Hepatitis Viruses

HEPATITIS TYPE	VIRUS PARTICLE, NM	MORPHOLOGY	GENOME ^a	CLASSIFICATION	ANTIGEN(s)	ANTIBODIES	REMARKS
HAV	27	Icosahedral nonenveloped	7.5-kb RNA, linear, ss, +	Hepatovirus	HAV	Anti-HAV	Early fecal shedding. Diagnosis: IgM anti-HAV Previous infection: IgG anti-HAV
HBV	42 27 22	Double-shelled virion (surface and core) spherical Nucleocapsid core Spherical and filamentous; represents excess virus coat material	3.2-kb DNA, circular, ss/ds	Hepadnavirus	HBsAg HBcAg HBeAg HBcAg HBeAg HBsAg	Anti-HBs Anti-HBc Anti-HBe Anti-HBc Anti-HBe Anti-HBs	Bloodborne virus; carrier state Acute diagnosis: HBsAg, IgM anti-HBc Chronic diagnosis: IgG anti-HBc, HBsAg Markers of replication: HBeAg, HBV DNA Liver, lymphocytes, other organs Nucleocapsid contains DNA and DNA polymerase; present in hepatocyte nucleus; HBcAg does not circulate; HBeAg (soluble, nonparticulate) and HBV DNA circulate—correlate with infectivity and complete virions

HEPATITIS TYPE	VIRUS PARTICLE, NM	MORPHOLOGY	GENOME ^a	CLASSIFICATION	ANTIGEN(s)	ANTIBODIES	REMARKS
							HBsAg detectable in >95% of patients with acute hepatitis B; found in serum, body fluids, hepatocyte cytoplasm; anti-HBs appears following infection—protective antibody
HCV	Approx. 50–80	Enveloped	9.4-kb RNA, linear, ss, +	Hepacivirus	HCV core antigen	Anti-HCV	Bloodborne agent, formerly labeled non-A, non-B hepatitis Acute diagnosis: anti-HCV, HCV RNA Chronic diagnosis: anti-HCV, HCV RNA; cytoplasmic location in hepatocytes

HEPATITIS TYPE	VIRUS PARTICLE, NM	MORPHOLOGY	GENOME ^a	CLASSIFICATION	ANTIGEN(s)	ANTIBODIES	REMARKS
HDV	35–37	Enveloped hybrid particle with HBsAg coat and HDV core	1.7-kb RNA, circular, ss, –	Resembles viroids and plant satellite viruses (genus Deltavirus)	HBsAg HDag	Anti-HBs Anti-HDV	Defective RNA virus, requires helper function of HBV (hepadnaviruses); HDV antigen (HDag) present in hepatocyte nucleus Diagnosis: anti-HDV, HDV RNA; HBV/HDV co-infection—IgM anti-HBc and anti-HDV; HDV superinfection—IgG anti-HBc and anti-HDV
HEV	32–34	Nonenveloped icosahedral	7.6-kb RNA, linear, ss, +	Hepevirus	HEV antigen	Anti-HEV	Agent of enterically transmitted hepatitis; rare in the United States; occurs in Asia, Mediterranean countries, Central America Diagnosis: IgM/IgG anti-HEV (assays not routinely available); virus in stool, bile, hepatocyte cytoplasm

^ass, single-strand; ss/ds, partially single-strand, partially double-strand; –, minus-strand; +, plus-strand.

Note: See text for abbreviations.

Envelope HBsAg subdeterminants include a common group-reactive antigen, *a*, shared by all HBsAg isolates and one of several subtype-specific antigens—*d* or *y*, *w* or *r*—as well as other specificities. Hepatitis B isolates fall into one of at least 8 subtypes and 10 genotypes (A–J). Geographic distribution of genotypes and subtypes varies; genotypes A (corresponding to subtype *adw*) and D (*ayw*) predominate in the United States and Europe, whereas genotypes B (*adw*) and C (*adr*) predominate in Asia. Clinical course and outcome are independent of subtype, but genotype B appears to be associated with less rapidly progressive liver disease and cirrhosis and a lower likelihood, or delayed appearance, of hepatocellular carcinoma than genotype C or D. Patients with genotype A are more likely to clear circulating viremia and achieve hepatitis B e antigen (HBeAg) and HBsAg seroconversion, both spontaneously and in response to antiviral therapy. In addition, “precore” mutations are favored by certain genotypes (see below).

Upstream of the S gene are the pre-S genes (Fig. 332-3), which code for pre-S gene products, including receptors on the HBV surface for polymerized human serum albumin and for hepatocyte membrane proteins. The pre-S region actually consists of both pre-S1 and pre-S2. Depending on where translation is initiated, three potential HBsAg gene products are synthesized. The protein product of the S gene is HBsAg (*major protein*), the product of the S region plus the adjacent pre-S2 region is the *middle protein*, and the product of the pre-S1 plus pre-S2 plus S regions is the *large protein*. Compared with the smaller spherical and tubular particles of HBV, complete 42-nm virions are enriched in the large protein. Both pre-S proteins and their respective antibodies can be detected during HBV infection, and the period of pre-S antigenemia appears to coincide with other markers of virus replication, as detailed below; however, pre-S proteins have little clinical relevance and are not included in routine serologic testing repertoires.

The intact 42-nm virion contains a 27-nm nucleocapsid core particle. Nucleocapsid proteins are coded for by the C gene. The antigen expressed on the surface of the nucleocapsid core is *hepatitis B core antigen* (HBcAg), and its corresponding antibody is anti-HBc. A third HBV antigen is *hepatitis B e antigen* (HBeAg), a soluble, nonparticulate, nucleocapsid protein that is immunologically distinct from intact HBcAg but is a product of the same C gene. The C gene has two initiation codons: a precore and a core region (Fig. 332-3). If translation is initiated at the precore region, the protein product is HBeAg, which has a signal peptide that binds it to the smooth endoplasmic reticulum, the secretory apparatus of the cell, leading to its secretion into the circulation. If translation begins at the core region, HBcAg is the protein product; it has no signal peptide and is not secreted, but it assembles into nucleocapsid particles, which bind to and incorporate RNA, and which, ultimately, contain HBV DNA. Also packaged within the nucleocapsid core is a DNA polymerase, which directs replication and repair of HBV DNA. When packaging within viral proteins is complete, synthesis of the incomplete plus strand stops; this accounts for the single-strand gap and for differences in the size of the gap. HBcAg particles remain in the hepatocyte, where they are readily detectable by immunohistochemical staining and are exported after encapsidation by an envelope of HBsAg. Therefore, naked core particles do not circulate in the serum. The secreted nucleocapsid protein, HBeAg, provides a convenient, readily detectable, qualitative marker of HBV replication and relative infectivity.

HBsAg-positive serum containing HBeAg is more likely to be highly infectious and to be associated with the presence of hepatitis B virions (and detectable HBV DNA, see below) than HBeAg-negative or anti-HBe-positive serum. For example, HBsAg-positive mothers who are HBeAg-positive almost invariably (>90%) transmit hepatitis B infection to their offspring, whereas HBsAg-positive mothers with anti-HBe rarely (10–15%) infect their offspring.

Early during the course of acute hepatitis B, HBeAg appears transiently; its disappearance may be a harbinger of clinical improvement and resolution of infection. Persistence of HBeAg in serum beyond the first 3 months of acute infection may be predictive of the development of chronic infection, and the presence of HBeAg during chronic hepatitis B tends to be associated with ongoing viral replication, infectivity, and inflammatory liver injury (except during the early decades after perinatally acquired HBV infection; see below).

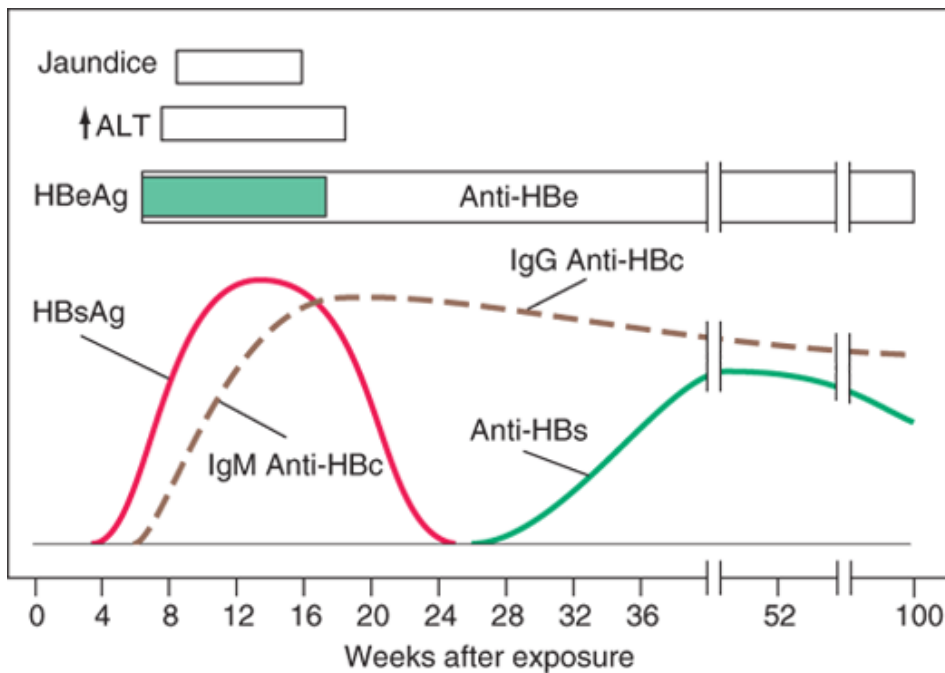
The third and largest of the HBV genes, the P gene (Fig. 332-3), codes for HBV DNA polymerase; as noted above, this enzyme has both DNA-dependent DNA polymerase and RNA-dependent reverse transcriptase activities. The fourth gene, X, codes for a small, nonparticulate protein, *hepatitis B x antigen* (HBxAg), that is capable of transactivating the transcription of both viral and cellular genes (Fig. 332-3). In the cytoplasm, HBxAg effects calcium release (possibly from mitochondria), which activates signal-transduction pathways that lead to stimulation of HBV reverse transcription and HBV DNA replication. Such transactivation may enhance the replication of HBV, leading to the clinical association observed between the expression of HBxAg and antibodies to it in patients with severe chronic hepatitis and hepatocellular carcinoma. The transactivating activity can enhance the transcription and replication of other viruses besides HBV, such as HIV. Cellular processes transactivated by X include the human interferon- γ gene and class I major histocompatibility genes; potentially, these effects could contribute to enhanced susceptibility of HBV-infected hepatocytes to cytolytic T cells. The expression of X can also induce programmed cell death (apoptosis). The clinical relevance of HBxAg is limited, however, and testing for it is not part of routine clinical practice.

SEROLOGIC AND VIROLOGIC MARKERS

After a person is infected with HBV, the first virologic marker detectable in serum within 1–12 weeks, usually between 8 and 12 weeks, is HBsAg (Fig. 332-4). Circulating HBsAg precedes elevations of serum aminotransferase activity and clinical symptoms by 2–6 weeks and remains detectable during the entire icteric or symptomatic phase of acute hepatitis B and beyond. In typical cases, HBsAg becomes undetectable 1–2 months after the onset of jaundice and rarely persists beyond 6 months. After HBsAg disappears, antibody to HBsAg (anti-HBs) becomes detectable in serum and remains detectable indefinitely thereafter. Because HBeAg is intracellular and, when in the serum, sequestered within an HBsAg coat, naked core particles do not circulate in serum, and therefore HBeAg is not detectable routinely in the serum of patients with HBV infection. By contrast, anti-HBe is readily demonstrable in serum, beginning within the first 1–2 weeks after the appearance of HBsAg and preceding detectable levels of anti-HBs by weeks to months. Because variability exists in the time of appearance of anti-HBs after HBV infection, occasionally a gap of several weeks or longer may separate the disappearance of HBsAg and the appearance of anti-HBs. During this “gap” or “window” period, anti-HBe may represent the only serologic evidence of current or recent HBV infection, and blood containing anti-HBe in the absence of HBsAg and anti-HBs has been implicated in transfusion-associated hepatitis B. In part because the sensitivity of immunoassays for HBsAg and anti-HBs has increased, however, this window period is rarely encountered. In some persons, years after HBV infection, anti-HBe may persist in the circulation longer than anti-HBs. Therefore, isolated anti-HBe does not necessarily indicate active virus replication; most instances of isolated anti-HBe represent hepatitis B infection in the remote past. Rarely, however, isolated anti-HBe represents low-level hepatitis B viremia, with HBsAg below the detection threshold, and, occasionally, isolated anti-HBe represents a cross-reacting or false-positive immunologic specificity. Recent and remote HBV infections can be distinguished by determination of the immunoglobulin class of anti-HBe. Anti-HBe of the IgM class (IgM anti-HBe) predominates during the first 6 months after acute infection, whereas IgG anti-HBe is the predominant class of anti-HBe beyond 6 months. Therefore, patients with current or recent acute hepatitis B, including those in the anti-HBe window, have IgM anti-HBe in their serum. In patients who have recovered from hepatitis B in the remote past as well as those with chronic HBV infection, anti-HBe is predominantly of the IgG class. Infrequently, in ≤ 1 –5% of patients with acute HBV infection, levels of HBsAg are too low to be detected; in such cases, the presence of IgM anti-HBe establishes the diagnosis of acute hepatitis B. When isolated anti-HBe occurs in the rare patient with chronic hepatitis B whose HBsAg level is below the sensitivity threshold of contemporary immunoassays (a low-level carrier), anti-HBe is of the IgG class. Generally, in persons who have recovered from hepatitis B, anti-HBs and anti-HBe persist indefinitely.

FIGURE 332-4

Scheme of typical clinical and laboratory features of acute hepatitis B. ALT, alanine aminotransferase.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

The temporal association between the appearance of anti-HBs and resolution of HBV infection as well as the observation that persons with anti-HBs in serum are protected against reinfection with HBV suggests that *anti-HBs is the protective antibody*. Therefore, strategies for prevention of HBV infection are based on providing susceptible persons with circulating anti-HBs (see below). Occasionally, in ~10% of patients with chronic hepatitis B, low-level, low-affinity anti-HBs can be detected. This antibody is directed against a subtype determinant different from that represented by the patient's HBsAg; its presence is thought to reflect the stimulation of a related clone of antibody-forming cells, but it has no clinical relevance and does not signal imminent clearance of hepatitis B. These patients with HBsAg and such nonneutralizing anti-HBs should be categorized as having chronic HBV infection.

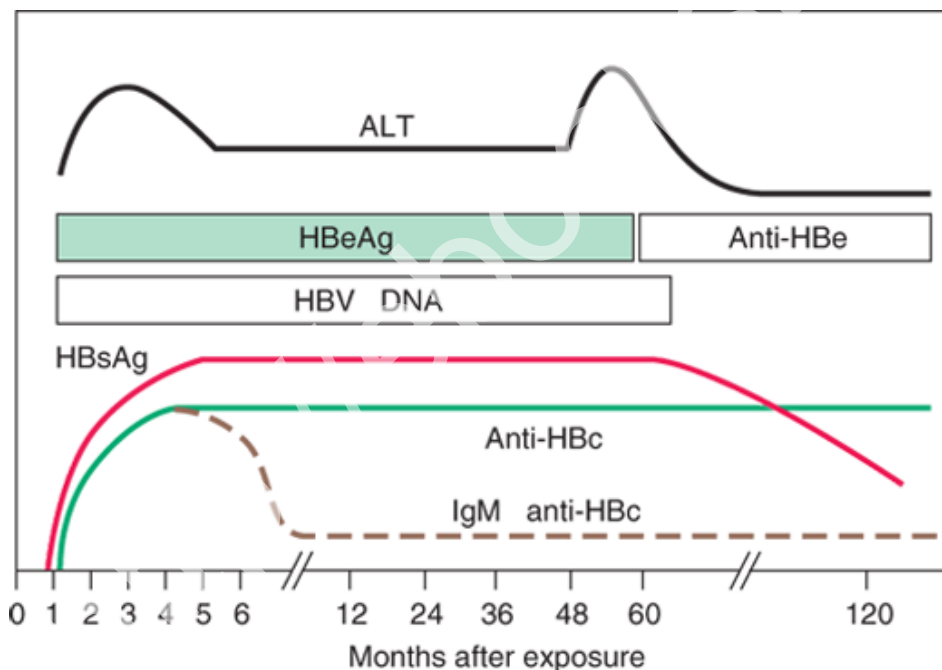
The other readily detectable serologic marker of HBV infection, HBeAg, appears concurrently with or shortly after HBsAg. Its appearance coincides temporally with high levels of virus replication and reflects the presence of circulating intact virions and detectable HBV DNA (with the notable exception of patients with precore mutations who cannot synthesize HBeAg—see “Molecular Variants”). Pre-S1 and pre-S2 proteins are also expressed during periods of peak replication, but assays for these gene products are not routinely available. In self-limited HBV infections, HBeAg becomes undetectable shortly after peak elevations in aminotransferase activity, before the disappearance of HBsAg, and anti-HBe then becomes detectable, coinciding with a period of relatively lower infectivity (Fig. 332-4). Because markers of HBV replication appear transiently during acute infection, testing for such markers is of little clinical utility in typical cases of acute HBV infection. In contrast, markers of HBV replication provide valuable information in patients with protracted infections.

Departing from the pattern typical of acute HBV infections, in chronic HBV infection, HBsAg remains detectable beyond 6 months, anti-HBc is primarily of the IgG class, and anti-HBs is either undetectable or detectable at low levels (see “Laboratory Features”) (Fig. 332-5). During early chronic HBV infection, HBV DNA can be detected both in serum and in hepatocyte nuclei, where it is present in free or episomal form. This relatively highly *replicative stage* of HBV infection is the time of maximal infectivity and liver injury; HBeAg is a qualitative marker and HBV DNA a quantitative marker of this replicative phase, during which all three forms of HBV circulate, including intact virions. Over time, the relatively replicative phase of chronic HBV infection gives way to a relatively *nonreplicative phase*. This occurs at a rate of ~10% per year and is accompanied by seroconversion from HBeAg to anti-HBe. In many cases, this seroconversion coincides

with a transient, usually mild, acute hepatitis-like elevation in aminotransferase activity, believed to reflect cell-mediated immune clearance of virus-infected hepatocytes. In the nonreplicative phase of chronic infection, when HBV DNA is demonstrable in hepatocyte nuclei, it tends to be integrated into the host genome. In this phase, only spherical and tubular forms of HBV, *not intact virions*, circulate, and liver injury tends to subside. Most such patients would be characterized as *inactive HBV carriers*. In reality, the designations *replicative* and *nonreplicative* are only relative; even in the so-called nonreplicative phase, HBV replication can be detected at levels of approximately $\leq 10^3$ virions/mL with highly sensitive amplification probes such as the polymerase chain reaction (PCR); below this replication threshold, liver injury and infectivity of HBV are limited to negligible. Still, the distinctions are pathophysiologically and clinically meaningful. Occasionally, nonreplicative HBV infection converts back to replicative infection. Such spontaneous reactivations are accompanied by reexpression of HBeAg and HBV DNA, and sometimes of IgM anti-HBc, as well as by exacerbations of liver injury. Because high-titer IgM anti-HBc can reappear during acute exacerbations of chronic hepatitis B, relying on IgM anti-HBc versus IgG anti-HBc to distinguish between acute and chronic hepatitis B infection, respectively, may not always be reliable; in such cases, patient history and additional follow-up monitoring over time are invaluable in helping to distinguish *de novo* acute hepatitis B infection from acute exacerbation of chronic hepatitis B infection.

FIGURE 332-5

Scheme of typical laboratory features of wild-type chronic hepatitis B. HBeAg and hepatitis B virus (HBV) DNA can be detected in serum during the relatively *replicative phase* of chronic infection, which is associated with infectivity and liver injury. Seroconversion from the replicative phase to the relatively *nonreplicative phase* occurs at a rate of ~10% per year and is heralded by an acute hepatitis-like elevation of alanine aminotransferase (ALT) activity; during the nonreplicative phase, infectivity and liver injury are limited. In HBeAg-negative chronic hepatitis B associated with mutations in the precore region of the HBV genome, replicative chronic hepatitis B occurs in the absence of HBeAg.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

MOLECULAR VARIANTS

Variation occurs throughout the HBV genome, and clinical isolates of HBV that do not express typical viral proteins have been attributed to mutations in individual or even multiple gene locations. For example, variants have been described that lack nucleocapsid proteins (commonly), envelope proteins (very rarely), or both. Two categories of naturally

occurring HBV variants have attracted the most attention. One of these was identified initially in Mediterranean countries among patients with severe chronic HBV infection and detectable HBV DNA but with anti-HBe instead of HBeAg. These patients were found to be infected with an HBV mutant that contained an alteration in the precore region rendering the virus incapable of encoding HBeAg. Although several potential mutation sites exist in the pre-C region, the region of the C gene necessary for the expression of HBeAg (see “[Virology and Etiology](#)”), the most commonly encountered in such patients is a single base substitution, from G to A in the second to last codon of the pre-C gene at nucleotide 1896. This substitution results in the replacement of the TGG tryptophan codon by a stop codon (TAG), which prevents the translation of HBeAg. Another mutation, in the core-promoter region, prevents transcription of the coding region for HBeAg and yields an HBeAg-negative phenotype. Patients with such mutations in the precore region and who are unable to secrete HBeAg may have severe liver disease that progresses more rapidly to cirrhosis, or alternatively, they are identified clinically later in the course of the natural history of chronic hepatitis B, when the disease is more advanced. Both “wild-type” HBV and precore-mutant HBV can coexist in the same patient, or mutant HBV may arise late during wild-type HBV infection. In addition, clusters of fulminant hepatitis B in Israel and Japan were attributed to common-source infection with a precore mutant. Fulminant hepatitis B in North America and western Europe, however, occurs in patients infected with wild-type HBV, in the absence of precore mutants, and both precore mutants and other mutations throughout the HBV genome occur commonly, even in patients with typical, self-limited, milder forms of HBV infection. HBeAg-negative chronic hepatitis with mutations in the precore region is now the most frequently encountered form of hepatitis B in Mediterranean countries and in Europe. In the United States, where HBV genotype A (less prone to G1896A mutation) is prevalent, precore-mutant HBV is much less common; however, as a result of immigration from Asia and Europe, the proportion of HBeAg-negative hepatitis B–infected individuals has increased in the United States, and they now represent ~30–40% of patients with chronic hepatitis B. Characteristic of such HBeAg-negative chronic hepatitis B are lower levels of HBV DNA (usually $\leq 10^5$ IU/mL) and one of several patterns of aminotransferase activity—persistent elevations, periodic fluctuations above the normal range, and periodic fluctuations between the normal and elevated range.

The second important category of HBV mutants consists of *escape mutants*, in which a single amino acid substitution, from glycine to arginine, occurs at position 145 of the immunodominant *a* determinant common to all HBsAg subtypes. This HBsAg alteration leads to a critical conformational change that results in a loss of neutralizing activity by anti-HBs. This specific HBV/*a* mutant has been observed in two situations, active and passive immunization, in which humoral immunologic pressure may favor evolutionary change (“escape”) in the virus—in a small number of hepatitis B vaccine recipients who acquired HBV infection despite the prior appearance of neutralizing anti-HBs and in HBV-infected liver transplant recipients treated with a high-potency human monoclonal anti-HBs preparation. Although such mutants have not been recognized frequently, their existence raises a concern that may complicate vaccination strategies and serologic diagnosis.

Different types of mutations emerge during antiviral therapy of chronic hepatitis B with nucleoside analogues; such “YMDD” and similar mutations in the polymerase motif of HBV are described in [Chap. 334](#).

EXTRAHEPATIC SITES

Hepatitis B antigens and HBV DNA have been identified in extrahepatic sites, including the lymph nodes, bone marrow, circulating lymphocytes, spleen, and pancreas. Although the virus does not appear to be associated with tissue injury in any of these extrahepatic sites, its presence in these “remote” reservoirs has been invoked (but is not necessary) to explain the recurrence of HBV infection after orthotopic liver transplantation. The clinical relevance of such extrahepatic HBV is limited.

Hepatitis D

The delta hepatitis agent, or HDV, the only member of the genus *Deltavirus*, is a defective RNA virus that co-infects with and requires the helper function of HBV (or other hepadnaviruses) for its replication and expression. Slightly smaller than HBV, HDV is a formalin-sensitive, 35- to 37-nm virus with a hybrid structure. Its nucleocapsid expresses HDV antigen (HDAg), which bears no antigenic homology with any of the HBV antigens, and contains the virus genome. The HDV core is “encapsidated” by an outer envelope of HBsAg, indistinguishable from that of HBV except in its relative compositions of major, middle, and large HBsAg component proteins. The genome is a small, 1700-nucleotide, circular, single-strand RNA of negative polarity that is nonhomologous with HBV DNA (except for a small area of the polymerase gene) but that has features and the rolling circle model of replication common to genomes of plant satellite viruses or viroids. HDV RNA contains many areas of internal complementarity; therefore, it can fold on itself by internal base pairing to form an unusual, very stable, rodlike structure that contains a very stable, self-cleaving and self-ligating ribozyme. HDV RNA requires host RNA polymerase II for its replication in the hepatocyte nucleus via RNA-directed RNA synthesis by transcription of genomic RNA to a complementary antigenomic (plus strand) RNA; the antigenomic RNA, in turn, serves as a template for subsequent genomic RNA synthesis effected by host RNA polymerase I. HDV RNA has only one open reading frame, and HDAg, a product of the antigenomic strand, is the only known HDV protein; HDAg exists in two forms: a small, 195-amino-acid species, which plays a role in facilitating HDV RNA replication, and a large, 214-amino-acid species, which appears to suppress replication but is required for assembly of the antigen into virions. HDV antigens have been shown to bind directly to RNA polymerase II, resulting in stimulation of transcription. Although complete hepatitis D virions and liver injury require the cooperative helper function of HBV, intracellular replication of HDV RNA can occur without HBV. Genomic heterogeneity among HDV isolates has been described; however, pathophysiologic and clinical consequences of this genetic diversity have not been recognized. The clinical spectrum of hepatitis D is common to all eight genotypes identified, the predominant of which is genotype 1.

HDV can either infect a person simultaneously with HBV (*co-infection*) or superinfect a person already infected with HBV (*superinfection*); when HDV infection is transmitted from a donor with one HBsAg subtype to an HBsAg-positive recipient with a different subtype, HDV assumes the HBsAg subtype of the recipient, rather than the donor. Because HDV relies absolutely on HBV, the duration of HDV infection is determined by the duration of (and cannot outlast) HBV infection. HDV replication tends to suppress HBV replication; therefore, patients with hepatitis D tend to have lower levels of HBV replication. HDV antigen is expressed primarily in hepatocyte nuclei and is occasionally detectable in serum. During acute HDV infection, anti-HDV of the IgM class predominates, and 30–40 days may elapse after symptoms appear before anti-HDV can be detected. In self-limited infection, anti-HDV is low-titer and transient, rarely remaining detectable beyond the clearance of HBsAg and HDV antigen. In chronic HDV infection, anti-HDV circulates in high titer, and both IgM and IgG anti-HDV can be detected. HDV antigen in the liver and HDV RNA in serum and liver can be detected during HDV replication.

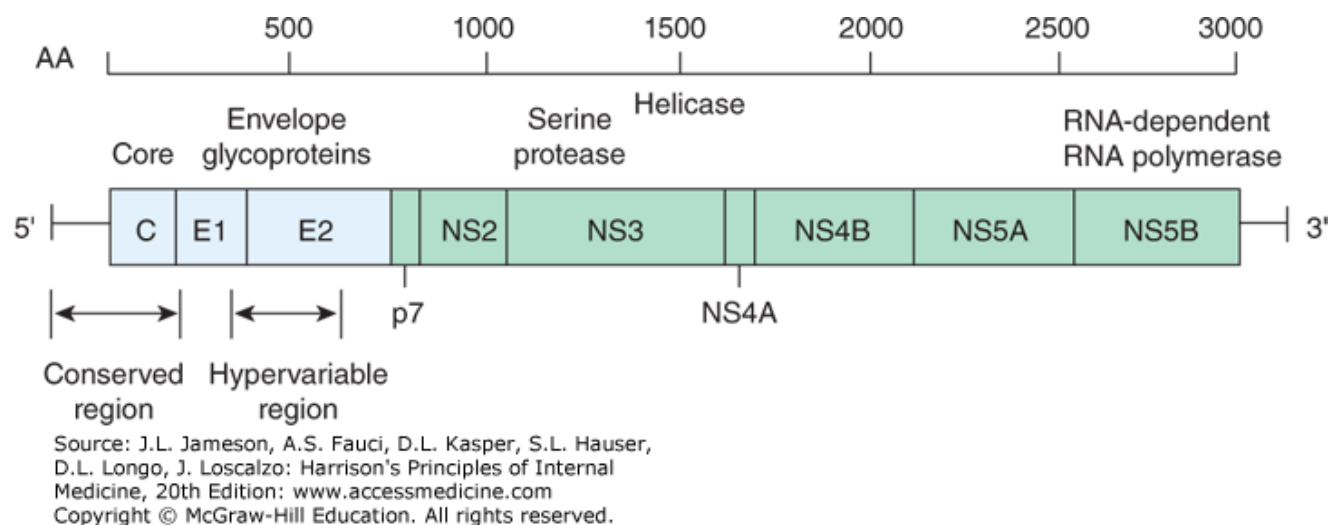
Hepatitis C

Hepatitis C virus, which, before its identification was labeled “non-A, non-B hepatitis,” is a linear, single-strand, positive-sense, 9600-nucleotide RNA virus, the genome of which is similar in organization to that of flaviviruses and pestiviruses; HCV is the only member of the genus *Hepacivirus* in the family Flaviviridae. The HCV genome contains a single, large open reading frame (ORF) (gene) that codes for a virus polyprotein of ~3000 amino acids, which is cleaved after translation to yield 10 viral proteins. The 5′ end of the genome consists of an untranslated region (containing an internal ribosomal entry site [IRES]) adjacent to the genes for three structural proteins, the nucleocapsid core protein, C, and two envelope glycoproteins, E1 and E2. The 5′ untranslated region and core gene are highly conserved among genotypes, but the envelope proteins are coded for by the hypervariable region, which varies from isolate to isolate and may allow the virus to evade host immunologic containment directed at accessible virus-envelope proteins. The 3′ end of the genome

also includes an untranslated region and contains the genes for seven nonstructural (NS) proteins: p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. p7 is a membrane ion channel protein necessary for efficient assembly and release of HCV. The NS2 **cysteine** protease cleaves NS3 from NS2, and the NS3-4A serine protease cleaves all the downstream proteins from the polyprotein. Important NS proteins involved in virus replication include the NS3 helicase; NS3-4A serine protease; the multifunctional membrane-associated phosphoprotein NS5A, an essential component of the viral replication membranous web (along with NS4B); and the NS5B RNA-dependent RNA polymerase (**Fig. 332-6**). Because HCV does not replicate via a DNA intermediate, it does not integrate into the host genome. Because HCV tends to circulate in relatively low titer, 10^3 – 10^7 virions/mL, visualization of the 50- to 80-nm virus particles remains difficult. Still, the replication rate of HCV is very high, 10^{12} virions per day; its half-life is 2.7 h. The chimpanzee is a helpful but cumbersome animal model. Although a robust, reproducible, small animal model is lacking, HCV replication has been documented in an immunodeficient mouse model containing explants of human liver and in transgenic mouse and rat models. Although in vitro replication is difficult, replicons in hepatocellular carcinoma–derived cell lines support replication of genetically manipulated, truncated, or full-length HCV RNA (but not intact virions); infectious pseudotyped retroviral HCV particles have been shown to yield functioning envelope proteins. In 2005, complete replication of HCV and intact 55-nm virions were described in cell culture systems. HCV entry into the hepatocyte occurs via the nonliver-specific CD81 receptor and the liver-specific tight junction protein claudin-1. A growing list of additional host receptors to which HCV binds on cell entry includes occludin, low-density lipoprotein receptors, glycosaminoglycans, scavenger receptor B1, and epidermal growth factor receptor, among others. Relying on the same assembly and secretion pathway as low-density and very-low-density lipoproteins, HCV is a lipovirion and masquerades as a lipoprotein, which may limit its visibility to the adaptive immune system and explain its ability to evade immune containment and clearance. After viral entry and uncoating, translation is initiated by the IRES on the endoplasmic reticulum membrane, and the HCV polyprotein is cleaved during translation and posttranslationally by host cellular proteases as well as HCV NS2-3 and NS3-4A proteases. Host cofactors involved in HCV replication include cyclophilin A, which binds to NS5A and yields conformational changes required for viral replication, and liver-specific host microRNA miR-122.

FIGURE 332-6

Organization of the hepatitis C virus genome and its associated, 3000-amino-acid (AA) proteins. The three structural genes at the 5' end are the core region, C, which codes for the nucleocapsid, and the envelope regions, E1 and E2, which code for envelope glycoproteins. The 5' untranslated region and the C region are highly conserved among isolates, whereas the envelope domain E2 contains the hypervariable region. At the 3' end are seven nonstructural (NS) regions—p7, a membrane protein adjacent to the structural proteins that appears to function as an ion channel; NS2, which codes for a **cysteine** protease; NS3, which codes for a serine protease and an RNA helicase; NS4 and NS4B; NS5A, a multifunctional membrane-associated phosphoprotein, an essential component of the viral replication membranous web; and NS5B, which codes for an RNA-dependent RNA polymerase. After translation of the entire polyprotein, individual proteins are cleaved by both host and viral proteases.

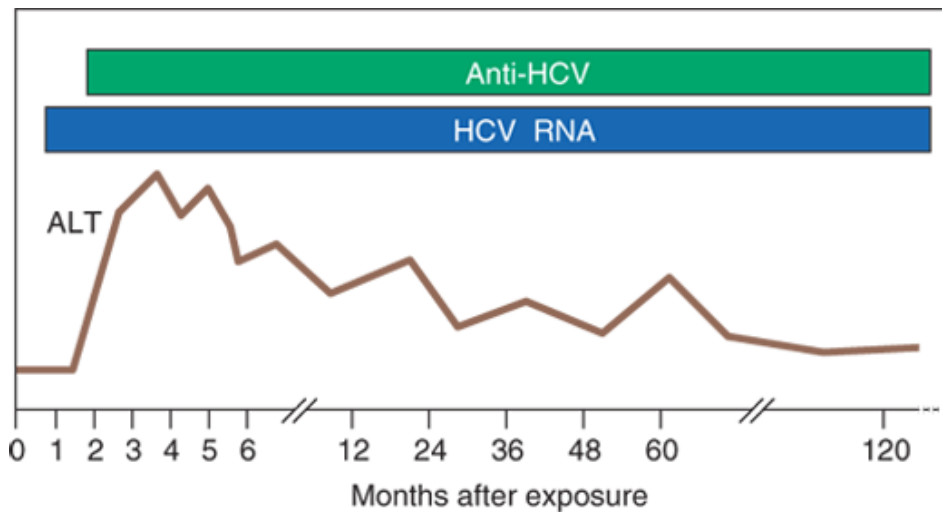


At least six distinct major genotypes (and a minor genotype 7), as well as >50 subtypes within genotypes, of HCV have been identified by nucleotide sequencing. Genotypes differ from one another in sequence homology by $\geq 30\%$, and subtypes differ by $\sim 20\%$. Because divergence of HCV isolates within a genotype or subtype and within the same host may vary insufficiently to define a distinct genotype, these intragenotypic differences are referred to as *quasispecies* and differ in sequence homology by only a few percent. The genotypic and quasispecies diversity of HCV, resulting from its high mutation rate, interferes with effective humoral immunity. Neutralizing antibodies to HCV have been demonstrated, but they tend to be short-lived, and HCV infection does not induce lasting immunity against reinfection with different virus isolates or even the same virus isolate. Thus, neither *heterologous* nor *homologous* immunity appears to develop commonly after acute HCV infection. Some HCV genotypes are distributed worldwide, whereas others are more geographically confined (see “[Epidemiology and Global Features](#)”). In addition, differences exist among genotypes in responsiveness to antiviral therapy but not in pathogenicity or clinical progression (except for genotype 3, in which hepatic steatosis and clinical progression are more likely).

Currently available, third-generation immunoassays, which incorporate proteins from the core, NS3, and NS5 regions, detect anti-HCV antibodies during acute infection. The most sensitive indicator of HCV infection is the presence of HCV RNA, which requires molecular amplification by PCR or transcription-mediated amplification (TMA) ([Fig. 332-7](#)). To allow standardization of the quantification of HCV RNA among laboratories and commercial assays, HCV RNA is reported as international units (IUs) per milliliter; quantitative assays with a broad dynamic range are available that allow detection of HCV RNA with a sensitivity as low as 5 IU/mL. HCV RNA can be detected within a few days of exposure to HCV—well before the appearance of anti-HCV—and tends to persist for the duration of HCV infection. Application of sensitive molecular probes for HCV RNA has revealed the presence of replicative HCV in peripheral blood lymphocytes of infected persons; however, as is the case for HBV in lymphocytes, the clinical relevance of HCV lymphocyte infection is not known.

FIGURE 332-7

Scheme of typical laboratory features during acute hepatitis C progressing to chronicity. Hepatitis C virus (HCV) RNA is the first detectable event, preceding alanine aminotransferase (ALT) elevation and the appearance of anti-HCV.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Hepatitis E

Previously labeled *epidemic or enterically transmitted non-A, non-B hepatitis*, HEV is an enterically transmitted virus that causes clinically apparent hepatitis primarily in India, Asia, Africa, and Central America; in those geographic areas, HEV is the most common cause of acute hepatitis; one-third of the global population appears to have been infected. This agent, with epidemiologic features resembling those of hepatitis A, is a 27- to 34-nm, nonenveloped, heat-stable, HAV-like virus with a 7200-nucleotide, single-strand, positive-sense RNA genome. HEV has three overlapping ORFs (genes), the largest of which, *ORF1*, encodes nonstructural proteins involved in virus replication (viral replicase). A middle-sized gene, *ORF2*, encodes the nucleocapsid protein, the major structural protein, and the smallest, *ORF3*, encodes a small structural protein involved in virus particle secretion. All HEV isolates appear to belong to a single serotype, despite genomic heterogeneity of up to 25% and the existence of five genotypes, only four of which have been detected in humans; genotypes 1 and 2 (common in developing countries) appear to be more virulent, whereas genotypes 3 (the most common in the United States and Europe) and 4 (seen in China) are more attenuated and account for subclinical infections. Contributing to the perpetuation of this virus are animal reservoirs, most notably in swine but also in camels, deer, rats, and rabbits, among others. No genomic or antigenic homology, however, exists between HEV and HAV or other picornaviruses; and HEV, although resembling caliciviruses, is sufficiently distinct from any known agent to merit its own classification as a unique genus, *Hepevirus*, within the family Hepeviridae. The virus has been detected in stool, bile, and liver, and is excreted in the stool during the late incubation period. Both IgM anti-HEV during early acute infection and IgG anti-HEV predominating after the first 3 months can be detected. The presence of HEV RNA in serum and stool accompanies acute infection; viremia resolves as clinical-biochemical recovery ensues, while HEV RNA in stool may outlast viremia by several weeks. Currently, availability and reliability of serologic/virologic testing for HEV infection is limited—and not FDA-approved or licensed—but can be done in specialized laboratories (e.g., the Centers for Disease Control and Prevention).

PATHOGENESIS

Under ordinary circumstances, none of the hepatitis viruses is known to be directly cytopathic to hepatocytes. Evidence suggests that the clinical manifestations and outcomes after acute liver injury associated with viral hepatitis are determined by the immunologic responses of the host. Among the viral hepatitises, the immunopathogenesis of hepatitis B and C has been studied most extensively.

Hepatitis B

For HBV, the existence of inactive hepatitis B carriers with normal liver histology and function suggests that the virus is not directly cytopathic. The fact that patients with defects in cellular immune competence are more likely to remain chronically infected rather than to clear HBV supports the role of cellular immune responses in the pathogenesis of hepatitis B-related liver injury. The model that has the most experimental support involves cytolytic T cells sensitized specifically to recognize host and hepatitis B viral antigens on the liver cell surface. Nucleocapsid proteins (HBcAg and possibly HBeAg), present on the cell membrane in minute quantities, are the viral target antigens that, with host antigens, invite cytolytic T cells to destroy HBV-infected hepatocytes. Differences in the robustness and broad polyclonality of CD8⁺ cytolytic T cell responsiveness; in the level of HBV-specific helper CD4⁺ T cells; in attenuation, depletion, and exhaustion of virus-specific T cells; in viral T cell epitope escape mutations that allow the virus to evade T cell containment; and in the elaboration of antiviral cytokines by T cells have been invoked to explain differences in outcomes between those who recover after acute hepatitis and those who progress to chronic hepatitis, or between those with mild and those with severe (fulminant) acute HBV infection.

Although a robust cytolytic T cell response occurs and eliminates virus-infected liver cells during acute hepatitis B, >90% of HBV DNA has been found in experimentally infected chimpanzees to disappear from the liver and blood before maximal T cell infiltration of the liver and before most of the biochemical and histologic evidence of liver injury. This observation suggests that components of the innate immune system and inflammatory cytokines, independent of cytopathic antiviral mechanisms, participate in the early immune response to HBV infection; this effect has been shown to represent elimination of HBV replicative intermediates from the cytoplasm and covalently closed circular viral DNA from the nucleus of infected hepatocytes. In turn, the innate immune response to HBV infection is mediated largely by natural killer (NK) cell cytotoxicity, activated by immunosuppressive cytokines (e.g., interleukin [IL] 10 and transforming growth factor [TGF] β), reduced signals from inhibitory receptor expression (e.g., major histocompatibility complex), or increased signals from activating receptor expression on infected hepatocytes. In addition, NK cells reduce helper CD4⁺ cells, which results in reduced CD8⁺ cells and exhaustion of the virus-specific T cell response to HBV infection. Ultimately, HBV-HLA-specific cytolytic T cell responses of the adaptive immune system are felt to be responsible for recovery from HBV infection.

Debate continues over the relative importance of viral and host factors in the pathogenesis of HBV-associated liver injury and its outcome. As noted above, precore genetic mutants of HBV have been associated with the more severe outcomes of HBV infection (severe chronic and fulminant hepatitis), suggesting that, under certain circumstances, relative pathogenicity is a property of the virus, not the host. The facts that concomitant HDV and HBV infections are associated with more severe liver injury than HBV infection alone and that cells transfected in vitro with the gene for HDV antigen express HDV antigen and then become necrotic in the absence of any immunologic influences are also consistent with a viral effect on pathogenicity. Similarly, in patients who undergo liver transplantation for end-stage chronic hepatitis B, occasionally, rapidly progressive liver injury appears in the new liver. This clinical pattern is associated with an unusual histologic pattern in the new liver, *fibrosing cholestatic hepatitis*, which, ultrastructurally, appears to represent a choking of the cell with overwhelming quantities of HBsAg. This observation suggests that, under the influence of the potent immunosuppressive agents required to prevent allograft rejection, HBV may have a direct cytopathic effect on liver cells, independent of the immune system.

Although the precise mechanism of liver injury in HBV infection remains elusive, studies of nucleocapsid proteins have shed light on the profound immunologic tolerance to HBV of babies born to mothers with highly replicative (HBeAg-positive), chronic HBV infection. In HBeAg-expressing transgenic mice, in utero exposure to HBeAg, which is sufficiently small to traverse the placenta, induces T cell tolerance to both nucleocapsid proteins. This, in turn, may explain why,

when infection occurs so early in life, immunologic clearance does not occur, and protracted, lifelong infection ensues. An alternative explanation proposed to explain why robust liver injury does not accompany neonatal HBV infection but predisposes to chronic infection is defective priming of HBV-specific T cells during in utero exposure to HBV.

An important distinction should be drawn between HBV infection acquired at birth, common in endemic areas, such as East Asia, and infection acquired in adulthood, common in the West. Infection in the neonatal period is associated with the acquisition of what appears to be a high level of immunologic tolerance to HBV and absence of an acute hepatitis illness, but the almost invariable establishment of chronic, often lifelong infection. Neonatally acquired HBV infection can culminate decades later in cirrhosis and hepatocellular carcinoma (see “[Complications and Sequelae](#)”). In contrast, when HBV infection is acquired during adolescence or early adulthood, the host immune response to HBV-infected hepatocytes tends to be robust, an acute hepatitis-like illness is the rule, and failure to recover is the exception. After adulthood-acquired infection, chronicity is uncommon, and the risk of hepatocellular carcinoma is very low. Based on these observations, some authorities categorize HBV infection into an “immunotolerant” phase, an “immunoreactive” phase, and an “inactive” phase. This somewhat simplistic formulation does not apply at all to the typical adult in the West with self-limited acute hepatitis B, in whom no period of immunologic tolerance occurs. Even among those with neonatally acquired HBV infection, in whom immunologic tolerance appears to be established definitively, immunologic responses to HBV infection have been demonstrated, and intermittent bursts of hepatic necroinflammatory activity punctuate the early decades of life during which liver injury appears to be quiescent (labeled by some as the “immunotolerant” phase; however, it more accurately is a period of dissociation between high-level HBV replication and a paucity of inflammatory liver injury). In addition, even when clinically apparent liver injury and progressive fibrosis emerge during later decades (the so-called immunoreactive, or immunointolerant, phase), the level of immunologic tolerance to HBV remains substantial. More accurately, in patients with neonatally acquired HBV infection, a dynamic equilibrium exists between tolerance and intolerance, the outcome of which determines the clinical expression of chronic infection. Persons infected as neonates tend to have a relatively higher level of immunologic tolerance (high replication, low necroinflammatory activity) during the early decades of life and a relatively lower level (but only rarely a loss) of tolerance (and necroinflammatory activity reflecting the level of virus replication) in the later decades of life.

Hepatitis C

Cell-mediated immune responses and elaboration by T cells of antiviral cytokines contribute to the multicellular innate and adaptive immune responses involved in the containment of infection and pathogenesis of liver injury associated with hepatitis C. The fact that HCV is so efficient in evading these immune mechanisms is a testament to its highly evolved ability to disrupt host immune responses at multiple levels. After exposure to HCV, the host cell identifies viral product motifs (pattern recognition receptors) that distinguish the virus from “self,” resulting in the elaboration of interferons and other cytokines that result in activation of innate and adaptive immune responses. Intrahepatic HLA class I-restricted cytolytic T cells directed at nucleocapsid, envelope, and nonstructural viral protein antigens have been demonstrated in patients with chronic hepatitis C; however, such virus-specific cytolytic T cell responses do not correlate adequately with the degree of liver injury or with recovery. Yet, a consensus has emerged supporting a role in the pathogenesis of HCV-associated liver injury of virus-activated CD4⁺ helper T cells that stimulate, via the cytokines they elaborate, HCV-specific CD8⁺ cytotoxic T cells. These responses appear to be more robust (higher in number, more diverse in viral antigen specificity, more functionally effective, and more long lasting) in those who recover from HCV infection than in those who have chronic infection. Contributing to chronic infection are a CD4⁺ proliferative defect that results in rapid contraction of CD4⁺ responses, mutations in CD8⁺ T cell-targeted viral epitopes that allow HCV to escape immune-mediated clearance, and upregulation of inhibitory receptors on functionally impaired, exhausted T cells. Although attention has focused on adaptive immunity, HCV proteins have been shown to interfere with innate immunity

by resulting in blocking of type 1 interferon responses and inhibition of interferon signaling and effector molecules in the interferon signaling cascade. Several HLA alleles have been linked with self-limited hepatitis C, the most convincing of which is the CC haplotype of the *IL28B* gene, which codes for interferon $\lambda 3$, a component of innate immune antiviral defense. The *IL28B* association is even stronger when combined with HLA class II *DQB1* *03:01. The link between non-CC *IL28B* polymorphisms and failure to clear HCV infection has been explained by a chromosome 19q13.13 frameshift variant upstream of *IL28B*, the ΔG polymorphism of which creates an ORF in a novel interferon gene (*IFN-14*) associated with impaired HCV clearance. Also shown to contribute to limiting HCV infection are NK cells of the innate immune system that function when HLA class I molecules required for successful adaptive immunity are underexpressed. Both peripheral cytotoxicity and intrahepatic NK cell cytotoxicity are dysfunctional in persistent HCV infection. Adding to the complexity of the immune response, HCV core, NS4B, and NS5B have been shown to suppress the immunoregulatory nuclear factor (NF)- κ B pathway, resulting in reduced antiapoptotic proteins and a resultant increased vulnerability to tumor necrosis factor (TNF) α -mediated cell death. Patients with hepatitis C and unfavorable (non-CC, associated with reduced HCV clearance) *IL28B* alleles have been shown to have depressed NK cell/innate immune function. Of note, the emergence of substantial viral quasispecies diversity and HCV sequence variation allow the virus to evade attempts by the host to contain HCV infection by both humoral and cellular immunity.

Finally, cross-reactivity between viral antigens (HCV NS3 and NS5A) and host autoantigens (cytochrome P450 2D6) has been invoked to explain the association between hepatitis C and a subset of patients with autoimmune hepatitis and antibodies to liver-kidney microsomal (LKM) antigen (anti-LKM) (**Chap. 334**).

EXTRAHEPATIC MANIFESTATIONS

Immune complex-mediated tissue damage appears to play a pathogenetic role in the extrahepatic manifestations of acute hepatitis B. The occasional prodromal serum sickness-like syndrome observed in acute hepatitis B appears to be related to the deposition in tissue blood vessel walls of HBsAg-anti-HBs circulating immune complexes, leading to activation of the complement system and depressed serum complement levels.

In patients with chronic hepatitis B, other types of immune-complex disease may be seen. Glomerulonephritis with the nephrotic syndrome is observed occasionally; HBsAg, immunoglobulin, and C3 deposition has been found in the glomerular basement membrane. Whereas generalized vasculitis (polyarteritis nodosa) develops in considerably <1% of patients with chronic HBV infection, 20–30% of patients with polyarteritis nodosa have HBsAg in serum (**Chap. 356**). In these patients, the affected small- and medium-size arterioles contain HBsAg, immunoglobulins, and complement components. Another extrahepatic manifestation of viral hepatitis, essential mixed cryoglobulinemia (EMC), was reported initially to be associated with hepatitis B. The disorder is characterized clinically by arthritis, cutaneous vasculitis (palpable purpura), and, occasionally, glomerulonephritis and serologically by the presence of circulating cryoprecipitable immune complexes of more than one immunoglobulin class (**Chaps. 308 and 356**). Many patients with this syndrome have chronic liver disease, but the association with HBV infection is limited; instead, a substantial proportion has chronic HCV infection, with circulating immune complexes containing HCV RNA. Immune-complex glomerulonephritis is another recognized extrahepatic manifestation of chronic hepatitis C. Immune-complex disorders have been linked, albeit rarely, with both hepatitis A and E. In hepatitis E, rare neurologic complications have been postulated to result from both immunologic mechanisms and/or direct CNS infection with the virus.

PATHOLOGY

The typical morphologic lesions of all types of viral hepatitis are similar and consist of panlobular infiltration with mononuclear cells, hepatic cell necrosis, hyperplasia of Kupffer cells, and variable degrees of cholestasis. Hepatic cell

regeneration is present, as evidenced by numerous mitotic figures, multinucleated cells, and “rosette” or “pseudoacinar” formation. The mononuclear infiltration consists primarily of small lymphocytes, although plasma cells and eosinophils occasionally are present. Liver cell damage consists of hepatic cell degeneration and necrosis, cell dropout, ballooning of cells, and acidophilic degeneration of hepatocytes (forming so-called Councilman or apoptotic bodies). Large hepatocytes with a ground-glass appearance of the cytoplasm may be seen in chronic but not in acute HBV infection; these cells contain HBsAg and can be identified histochemically with orcein or aldehyde fuchsin. In uncomplicated viral hepatitis, the reticulin framework is preserved.

In hepatitis C, the histologic lesion is often remarkable for a relative paucity of inflammation, a marked increase in activation of sinusoidal lining cells, lymphoid aggregates, the presence of fat (more frequent in genotype 3 and linked to increased fibrosis), and, occasionally, bile duct lesions in which biliary epithelial cells appear to be piled up without interruption of the basement membrane. Occasionally, microvesicular steatosis occurs in hepatitis D. In hepatitis E, a common histologic feature is marked cholestasis. A cholestatic variant of slowly resolving acute hepatitis A also has been described.

A more severe histologic lesion, *bridging hepatic necrosis*, also termed *subacute* or *confluent necrosis* or *interface hepatitis*, is observed occasionally in acute hepatitis. “Bridging” between lobules results from large areas of hepatic cell dropout, with collapse of the reticulin framework. Characteristically, the bridge consists of condensed reticulum, inflammatory debris, and degenerating liver cells that span adjacent portal areas, portal to central veins, or central vein to central vein. This lesion had been thought to have prognostic significance; in many of the originally described patients with this lesion, a subacute course terminated in death within several weeks to months, or severe chronic hepatitis and cirrhosis developed; however, the association between bridging necrosis and a poor prognosis in patients with acute hepatitis has not been upheld. Therefore, although demonstration of this lesion in patients with chronic hepatitis has prognostic significance ([Chap. 334](#)), its demonstration during acute hepatitis is less meaningful, and liver biopsies to identify this lesion are no longer undertaken routinely in patients with acute hepatitis. In *massive hepatic necrosis* (fulminant hepatitis, “acute yellow atrophy”), the striking feature at postmortem examination is the finding of a small, shrunken, soft liver. Histologic examination reveals massive necrosis and dropout of liver cells of most lobules with extensive collapse and condensation of the reticulin framework. When histologic documentation is required in the management of fulminant or very severe hepatitis, a biopsy can be done by the angiographically guided transjugular route, which permits the performance of this invasive procedure in the presence of severe coagulopathy.

Immunohistochemical and electron-microscopic studies have localized HBsAg to the cytoplasm and plasma membrane of infected liver cells. In contrast, HBcAg predominates in the nucleus, but, occasionally, scant amounts are also seen in the cytoplasm and on the cell membrane. HDV antigen is localized to the hepatocyte nucleus, whereas HAV, HCV, and HEV antigens are localized to the cytoplasm.

EPIDEMIOLOGY AND GLOBAL FEATURES



Before the availability of serologic tests for hepatitis viruses, all viral hepatitis cases were labeled either as “infectious” or “serum” hepatitis. Modes of transmission overlap, however, and *a clear distinction among the different types of viral hepatitis cannot be made solely on the basis of clinical or epidemiologic features* ([Table 332-2](#)). The most accurate means to distinguish the various types of viral hepatitis involves specific serologic testing.

TABLE 332-2

Clinical and Epidemiologic Features of Viral Hepatitis

FEATURE	HAV	HBV	HCV	HDV	HEV
Incubation (days)	15–45, mean 30	30–180, mean 60–90	15–160, mean 50	30–180, mean 60–90	14–60, mean 40
Onset	Acute	Insidious or acute	Insidious or acute	Insidious or acute	Acute
Age preference	Children, young adults	Young adults (sexual and percutaneous), babies, toddlers	Any age, but more common in adults	Any age (similar to HBV)	Epidemic cases: young adults (20–40 years); sporadic cases: older adults (>60)
Transmission	+++	–	–	–	+++
Fecal-oral	Unusual	+++	+++	+++	–
	–	+++	± ^a	+	–
Percutaneous	±	++	± ^a	++	–
Perinatal					
Sexual					
Clinical	Mild	Occasionally severe	Moderate	Occasionally severe	Mild
Severity	0.1%		0.1%		1–2% ^e
Fulminant	None	0.1–1%	Common (85%)	5–20% ^b	None ^f
	None	Occasional (1–10%) (90% of neonates)	1.5–3.2%	Common ^d	None
Progression to chronicity	None		+	Variable ^g	None
Carrier	Excellent	0.1–30% ^c	Moderate	±	Good
Cancer		+ (neonatal infection)		Acute, good	
Prognosis		Worse with age, debility		Chronic, poor	
Prophylaxis	Ig, inactivated vaccine	HBIG, recombinant vaccine	None	HBV vaccine (none for HBV carriers)	Vaccine

FEATURE	HAV	HBV	HCV	HDV	HEV
Therapy	None	Interferon Lamivudine Adefovir Pegylated interferon ^h Entecavir ^h Telbivudine Tenofovir ^h	Pegylated interferon ribavirin telaprevir, ⁱ boceprevir, ⁱ simeprevir, sofosbuvir , ledipasvir, paritaprevir/ritonavir ombitasvir, dasabuvir daclatasvir, velpatasvir, grazoprevir, elbasvir	Pegylated interferon ±	None ^j

^aPrimarily with HIV co-infection and high-level viremia in index case; more likely in persons with multiple sex partners or sexually transmitted diseases; risk ~5%. ^bUp to 5% in acute HBV/HDV co-infection; up to 20% in HDV superinfection of chronic HBV infection. ^cVaries considerably throughout the world and in subpopulations within countries; see text. ^dIn acute HBV/HDV co-infection, the frequency of chronicity is the same as that for HBV; in HDV superinfection, chronicity is invariable. ^e10–20% in pregnant women. ^fExcept as observed in immunosuppressed liver allograft recipients or other immunosuppressed hosts. ^gCommon in Mediterranean countries; rare in North America and western Europe. ^hFirst-line agents. ⁱNo longer recommended. ^jAnecdotal reports and retrospective studies suggest that pegylated interferon and/or **ribavirin** are effective in treating chronic hepatitis E, observed in immunocompromised persons; **ribavirin** monotherapy has been used successfully in acute, severe hepatitis E.

Abbreviation: HBIG, hepatitis B immunoglobulin. See text for other abbreviations.

Hepatitis A

This agent is transmitted almost exclusively by the fecal-oral route. Person-to-person spread of HAV is enhanced by poor personal hygiene and overcrowding; large outbreaks as well as sporadic cases have been traced to contaminated food, water, milk, frozen raspberries and strawberries, green onions imported from Mexico, and shellfish (e.g., scallops imported from the Philippines used to make sushi, the culprit identified in a 2016 Hawaiian outbreak). Intrafamily and intrainstitutional spreads are also common. Early epidemiologic observations supported a predilection for hepatitis A to occur in late fall and early winter. In temperate zones, epidemic waves have been recorded every 5–20 years as new segments of nonimmune population appeared; however, in developed countries, the incidence of hepatitis A has been declining, presumably as a function of improved sanitation, and these cyclic patterns are no longer observed. No HAV carrier state has been identified after acute hepatitis A; perpetuation of the virus in nature depends presumably on nonepidemic, inapparent subclinical infection, ingestion of contaminated food or water in, or imported from, endemic areas, and/or contamination linked to environmental reservoirs.

In the general population, anti-HAV, a marker for previous HAV infection, increases in prevalence as a function of increasing age and of decreasing socioeconomic status. In the 1970s, serologic evidence of prior hepatitis A infection occurred in ~40% of urban populations in the United States, most of whose members never recalled having had a symptomatic case of hepatitis. In subsequent decades, however, the prevalence of anti-HAV has been declining in the United States. In developing countries, exposure, infection, and subsequent immunity are almost universal in childhood. As the frequency of subclinical childhood infections declines in developed countries, a susceptible cohort of adults

emerges. Hepatitis A tends to be more symptomatic in adults; therefore, paradoxically, as the frequency of HAV infection declines, the likelihood of clinically apparent, even severe, HAV illnesses increases in the susceptible adult population. Travel to endemic areas is a common source of infection for adults from nonendemic areas. More recently recognized epidemiologic foci of HAV infection include child care centers, neonatal intensive care units, promiscuous men who have sex with men, injection drug users, and unvaccinated close contacts of newly arrived international adopted children, most of whom emanate from countries with intermediate-to-high hepatitis A endemicity. Although hepatitis A is rarely bloodborne, several outbreaks have been recognized in recipients of clotting-factor concentrates. In the United States, the introduction of hepatitis A vaccination programs among children from high-incidence states has resulted in a >70% reduction in the annual incidence of new HAV infections and has shifted the burden of new infections from children to adults. In the 2007–2012 U.S. Public Health Service National Health and Nutrition Examination Survey (NHANES), the prevalence of anti-HAV in the U.S. population aged ≥ 20 years had declined to 24.2% from the 29.5% measured in NHANES 1999–2006. While universal childhood vaccination accounted for a high prevalence of vaccine-induced immunity in children aged 2–19 years, the lowest age-specific prevalence of anti-HAV (16.1–17.6%) occurred in adults in the fourth and fifth decades, respectively (aged 30–49 years). This is a subgroup of the population who remain susceptible to acute hepatitis A acquired during travel to endemic areas and from contaminated foods, especially those imported from endemic countries.

Hepatitis B

Percutaneous inoculation has long been recognized as a major route of hepatitis B transmission, but the outmoded designation “serum hepatitis” is an inaccurate label for the epidemiologic spectrum of HBV infection. As detailed below, most of the hepatitis transmitted by blood transfusion is not caused by HBV; moreover, in approximately two-thirds of patients with acute type B hepatitis, no history of an identifiable percutaneous exposure can be elicited. We now recognize that many cases of hepatitis B result from less obvious modes of nonpercutaneous or covert percutaneous transmission. HBsAg has been identified in almost every body fluid from infected persons, and at least some of these body fluids—most notably semen and saliva—are infectious, albeit less so than serum, when administered percutaneously or nonpercutaneously to experimental animals. Among the nonpercutaneous modes of HBV transmission, oral ingestion has been documented as a potential but inefficient route of exposure. By contrast, the two nonpercutaneous routes considered to have the greatest impact are intimate (especially sexual) contact and perinatal transmission.

In sub-Saharan Africa, intimate contact among toddlers is considered instrumental in contributing to the maintenance of the high frequency of hepatitis B in the population. Perinatal transmission occurs primarily in infants born to mothers with chronic hepatitis B or (rarely) mothers with acute hepatitis B during the third trimester of pregnancy or during the early postpartum period. Perinatal transmission is uncommon in North America and western Europe but occurs with great frequency and is the most important mode of HBV perpetuation in East Asia and developing countries. Although the precise mode of perinatal transmission is unknown, and although ~10% of infections may be acquired in utero, epidemiologic evidence suggests that most infections occur approximately at the time of delivery and are not related to breast-feeding (which is not contraindicated in women with hepatitis B). The likelihood of perinatal transmission of HBV correlates with the presence of HBeAg and high-level viral replication; 90% of HBeAg-positive mothers but only 10–15% of anti-HBe-positive mothers transmit HBV infection to their offspring. In most cases, acute infection in the neonate is clinically asymptomatic, but the child is very likely to remain chronically infected.

The >350–400 million persons with chronic HBV infection in the world constitute the main reservoir of hepatitis B in human beings. Whereas serum HBsAg is infrequent (0.1–0.5%) in normal populations in the United States and western Europe, a prevalence of up to 5–20% has been found in East Asia and in some tropical countries; in persons with Down’s

syndrome, lepromatous leprosy, leukemia, Hodgkin's disease, or polyarteritis nodosa; in patients with chronic renal disease on hemodialysis; and in injection drug users.

Other groups with high rates of HBV infection include spouses of acutely infected persons; sexually promiscuous persons (especially promiscuous men who have sex with men); health care workers exposed to blood; persons who require repeated transfusions especially with pooled blood-product concentrates (e.g., hemophiliacs); residents and staff of custodial institutions for the developmentally handicapped; prisoners; and, to a lesser extent, family members of chronically infected patients. In volunteer blood donors, the prevalence of anti-HBs, a reflection of previous HBV infection, ranges from 5% to 10%, but the prevalence is higher in lower socioeconomic strata, older age groups, and persons—including those mentioned above—exposed to blood products. Because of highly sensitive virologic screening of donor blood, the risk of acquiring HBV infection from a blood transfusion is 1 in 230,000.

Prevalence of infection, modes of transmission, and human behavior conspire to mold geographically different epidemiologic patterns of HBV infection. In East Asia and Africa, hepatitis B, a disease of the newborn and young children, is perpetuated by a cycle of maternal-neonatal spread. In North America and western Europe, hepatitis B is primarily a disease of adolescence and early adulthood, the time of life when intimate sexual contact and recreational and occupational percutaneous exposures tend to occur. To some degree, however, this dichotomy between high-prevalence and low-prevalence geographic regions has been minimized by immigration from high-prevalence to low-prevalence areas. For example, in the United States, NHANES data from 2007 to 2012 revealed an overall prevalence of current HBV infection (detectable HBsAg) of 0.3%; however, the prevalence in Asian persons, 93% of whom were foreign-born, was 10-fold higher, 3.1%, representing 50% of the U.S. national disease burden. The introduction of hepatitis B vaccine in the early 1980s and adoption of universal childhood vaccination policies in many countries resulted in a dramatic, ~90% decline in the incidence of new HBV infections in those countries as well as in the dire consequences of chronic infection, including hepatocellular carcinoma. In the United States, as demonstrated in NHANES 2007–2012, following the 1991 implementation of universal childhood vaccination, HBsAg seropositivity had declined in children aged 6–19 years to as low as 0.03%, an ~85% reduction. Populations and groups for whom HBV infection screening is recommended are listed in [Table 332-3](#).

TABLE 332-3

High-Risk Populations for Whom Hbv Infection Screening is Recommended

Persons born in countries/regions with a high ($\geq 8\%$) and intermediate ($\geq 2\%$) prevalence of HBV infection including immigrants and adopted children and including persons born in the United States who were not vaccinated as infants and whose parents emigrated from areas of high HBV endemicity

Household and sexual contacts of persons with hepatitis B

Babies born to HBsAg-positive mothers

Persons who have used injection drugs

Persons with multiple sexual contacts or a history of sexually transmitted disease

Men who have sex with men

Inmates of correctional facilities

Persons with elevated alanine or aspartate aminotransferase levels

Blood/plasma/organ/tissue/semen donors

Persons with HCV or HIV infection

Hemodialysis patients

Pregnant women

Persons who are the source of blood or body fluids that would be an indication for postexposure prophylaxis (e.g., needlestick, mucosal exposure, sexual assault)

Persons who require immunosuppressive or cytotoxic therapy (including anti-tumor necrosis factor α therapy for rheumatologic or inflammatory bowel disorders)

Hepatitis D

Infection with HDV has a worldwide distribution, but two epidemiologic patterns exist. In Mediterranean countries (northern Africa, southern Europe, the Middle East), HDV infection is endemic among those with hepatitis B, and the disease is transmitted predominantly by nonpercutaneous means, especially close personal contact. In nonendemic areas, such as the United States (where hepatitis D is rare among persons with chronic hepatitis B) and northern Europe, HDV infection is confined to persons exposed frequently to blood and blood products, primarily injection drug users, (especially so in HIV-infected injection drug users) and hemophiliacs. In the United States, the prevalence of HDV infection in the national population is 0.02% (NHANES 1999–2012); however, among HBsAg-positive persons, the prevalence of HDV infection is highest in injection drug users (11%) and hemophiliacs (19%). HDV infection can be introduced into a population through drug users or by migration of persons from endemic to nonendemic areas. Thus, patterns of population migration and human behavior facilitating percutaneous contact play important roles in the introduction and amplification of HDV infection. Occasionally, the migrating epidemiology of hepatitis D is expressed in explosive outbreaks of severe hepatitis, such as those that have occurred in remote South American villages (e.g., “Lábrea fever” in the Amazon basin) as well as in urban centers in the United States. Ultimately, such outbreaks of hepatitis D—either of co-infections with acute hepatitis B or of superinfections in those already infected with HBV—may blur the distinctions between endemic and nonendemic areas. On a global scale, HDV infection declined at the end of the 1990s. Even in Italy, an HDV-endemic area, public health measures introduced to control HBV infection (e.g., mass hepatitis B vaccination) resulted during the 1990s in a 1.5%/year reduction in the prevalence of HDV infection. Still, the frequency of HDV infection during the first decade of the twenty-first century has not fallen below levels reached during

the 1990s; the reservoir has been sustained by survivors infected during 1970–1980 and recent immigrants from still-endemic (e.g., Eastern Europe and Central Asia) to less-endemic countries.

Hepatitis C

Routine screening of blood donors for HBsAg and the elimination of commercial blood sources in the early 1970s reduced the frequency of, but did not eliminate, transfusion-associated hepatitis. During the 1970s, the likelihood of acquiring hepatitis after transfusion of voluntarily donated, HBsAg-screened blood was ~10% per patient (up to 0.9% per unit transfused); 90–95% of these cases were classified, based on serologic exclusion of hepatitis A and B, as “non-A, non-B” hepatitis. For patients requiring transfusion of pooled products, such as clotting factor concentrates, the risk was even higher, up to 20–30%.

During the 1980s, voluntary self-exclusion of blood donors with risk factors for AIDS and then the introduction of donor screening for anti-HIV reduced further the likelihood of transfusion-associated hepatitis to <5%. During the late 1980s and early 1990s, the introduction first of “surrogate” screening tests for non-A, non-B hepatitis (alanine aminotransferase [ALT] and anti-HBc, both shown to identify blood donors with a higher likelihood of transmitting non-A, non-B hepatitis to recipients) and, subsequently, after the discovery of HCV, first-generation immunoassays for anti-HCV reduced the frequency of transfusion-associated hepatitis even further. A prospective analysis of transfusion-associated hepatitis conducted between 1986 and 1990 showed that the frequency of transfusion-associated hepatitis at one urban university hospital fell from a baseline of 3.8% per patient (0.45% per unit transfused) to 1.5% per patient (0.19% per unit) after the introduction of surrogate testing and to 0.6% per patient (0.03% per unit) after the introduction of first-generation anti-HCV assays. The introduction of second-generation anti-HCV assays reduced the frequency of transfusion-associated hepatitis C to almost imperceptible levels—1 in 100,000—and these gains were reinforced by the application of third-generation anti-HCV assays and of automated PCR testing of donated blood for HCV RNA, which has resulted in a reduction in the risk of transfusion-associated HCV infection to 1 in 2.3 million transfusions.

In addition to being transmitted by transfusion, hepatitis C can be transmitted by other percutaneous routes, such as injection drug use. In addition, this virus can be transmitted by occupational exposure to blood, and the likelihood of infection is increased in hemodialysis units. Although the frequency of transfusion-associated hepatitis C fell as a result of blood-donor screening, the overall frequency of hepatitis C remained the same until the early 1990s, when the overall frequency of reported cases fell by 80%, in parallel with a reduction in the number of new cases in injection drug users. After the exclusion of anti-HCV-positive plasma units from the donor pool, rare, sporadic instances have occurred of hepatitis C among recipients of immunoglobulin preparations for intravenous (but not intramuscular) use.

Serologic evidence for HCV infection occurs in 90% of patients with a history of transfusion-associated hepatitis (almost all occurring before 1992, when second-generation HCV screening tests were introduced); hemophiliacs and others treated with clotting factors; injection drug users; 60–70% of patients with sporadic “non-A, non-B” hepatitis who lack identifiable risk factors; 0.5% of volunteer blood donors; and, in the NHANES survey conducted in the United States between 1999 and 2002, 1.6% of the general population in the United States, which translates into 4.1 million persons (3.2 million with viremia), the majority of whom are unaware of their infections. Moreover, such population surveys do not include higher-risk groups such as incarcerated persons, homeless persons, and active injection drug users, indicating that the actual prevalence is even higher. Comparable frequencies of HCV infection occur in most countries around the world, with 170 million persons infected worldwide, but extraordinarily high prevalences of HCV infection occur in certain countries such as Egypt, where >20% of the population (as high as 50% in persons born prior to 1960) in some cities is infected. The high frequency in Egypt is attributable to contaminated equipment used for medical procedures and unsafe injection practices in the 1950s to 1980s (during a campaign to eradicate schistosomiasis with

intravenous tartar emetic). In the United States, African Americans and Mexican Americans have higher frequencies of HCV infection than whites. Data from NHANES showed that between 1988 and 1994, 30- to 40-year-old men had the highest prevalence of HCV infection; however, in a survey conducted between 1999 and 2002, the peak age decile had shifted to those age 40–49 years; an increase in hepatitis C–related mortality has paralleled this secular trend, increasing since 1995 predominantly in the 45- to 65-year age group. Thus, despite an 80% reduction in new HCV infections during the 1990s, the prevalence of HCV infection in the population was sustained by an aging cohort that had acquired their infections three to four decades earlier, during the 1960s and 1970s, as a result predominantly of self-inoculation with recreational drugs. Retrospective phylogenetic mapping of >45,000 HCV genotype 1a isolates revealed that the hepatitis C epidemic emerged in the United States between 1940 and 1965, peaking in 1950 and aligning temporally with the post-World-War-II expansion of medical procedures (including re-use of glass syringes). Thus, HCV was amplified iatrogenically not only in Egypt but also in the United States; in the United States, the seeds sown by medical procedures in the 1950s were reaped in the 1960s and 1970s among transfusion recipients and injection drug users, even those whose drug use was confined to brief adolescent experimentation.

In NHANES 2003–2010, the prevalence of HCV infection (HCV RNA reactivity) in the United States had actually fallen to 1% (2.7 million persons) from 1.3% (3.2 million) the decade before (NHANES 1999–2002), attributable to deaths among the HCV-infected population. As death resulting from HIV infection fell after 1999, age-adjusted mortality associated with HCV infection surpassed that of HIV infection in 2007; >70% of HCV-associated deaths occurred in the “baby boomer” cohort born between 1945 and 1965. By 2012, HCV mortality had surpassed deaths from HIV, tuberculosis, hepatitis B, and 57 other notifiable infectious diseases (i.e., *all* infectious diseases) reported to the Centers for Disease Control and Prevention. In NHANES 1999–2002, compared to the 1.6% prevalence of HCV infection in the population at large, the prevalence in the 1945–1965 birth cohort was 3.2%, representing three-quarters of all infected persons. Therefore, in 2012, the Centers for Disease Control and Prevention recommended that all persons born between 1945 and 1965 be screened for hepatitis C, without ascertainment of risk, a recommendation shown to be cost-effective and predicted to identify 800,000 infected persons. Because of the availability of highly effective antiviral therapy, such screening would have the potential to avert 200,000 cases of cirrhosis and 47,000 cases of hepatocellular carcinoma and to prevent 120,000 hepatitis-related deaths; with the availability of the new generation of direct-acting antivirals (efficacy >95%, see [Chap 334](#)), screening baby boomers and treating those with hepatitis C have been predicted to reduce the HCV-associated disease burden by 50–70% through 2050.

Hepatitis C accounts for 40% of chronic liver disease, is the most frequent indication for liver transplantation, and is estimated to account for 8000–10,000 deaths per year in the United States. The distribution of HCV genotypes varies in different parts of the world. Worldwide, genotype 1 is the most common. In the United States, genotype 1 accounts for 70% of HCV infections, whereas genotypes 2 and 3 account for the remaining 30%; among African Americans, the frequency of genotype 1 is even higher (i.e., 90%). Genotype 4 predominates in Egypt; genotype 5 is localized to South Africa, genotype 6 to Hong Kong, and genotype 7 to Central Africa. Most asymptomatic blood donors found to have anti-HCV and ~20–30% of persons with reported cases of acute hepatitis C do not fall into a recognized risk group; however, many such blood donors do recall risk-associated behaviors when questioned carefully.

As a bloodborne infection, HCV potentially can be transmitted sexually and perinatally; however, both of these modes of transmission are inefficient for hepatitis C. Although 10–15% of patients with acute hepatitis C report having potential sexual sources of infection, most studies have failed to identify sexual transmission of this agent. The chances of sexual and perinatal transmission have been estimated to be ~5% but shown in a prospective study to be only 1% between monogamous sexual partners, well below comparable rates for HIV and HBV infections. Moreover, sexual transmission appears to be confined to such subgroups as persons with multiple sexual partners and sexually transmitted diseases; for example, isolated clusters of sexually transmitted HCV infection have been reported in HIV-infected men who have

sex with men. Breast-feeding does not increase the risk of HCV infection between an infected mother and her infant. Infection of health workers is not dramatically higher than among the general population; however, health workers are more likely to acquire HCV infection through accidental needle punctures, the efficiency of which is ~3%. Infection of household contacts is rare as well.

Besides persons born between 1945 and 1965, other groups with an increased frequency of HCV infection are listed in **Table 332-4**. In immunosuppressed individuals, levels of anti-HCV may be undetectable, and a diagnosis may require testing for HCV RNA. Although new acute cases of hepatitis C are rare outside of the injection-drug using community, newly diagnosed cases are common among otherwise healthy persons who experimented briefly with injection drugs, as noted above, three or four decades earlier. Such instances usually remain unrecognized for years, until unearthed by laboratory screening for routine medical examinations, insurance applications, and attempted blood donation. Although, overall, the annual incidence of new HCV infections has continued to fall, the rate of new infections has been increasing since 2002, amplified by the recent epidemic of opioid use, in a new cohort of young injection drug users, age 15–24 years (accounting for more than two-thirds of all acute cases), who, unlike older cohorts, had not learned to take precautions to prevent bloodborne infections.

TABLE 332-4

High-Risk Populations for Whom Hcv-Infection Screening is Recommended

Persons born between 1945 and 1965
 Persons who have ever used injection drugs
 Persons with HIV infection
 Hemophiliacs treated with clotting factor concentrates prior to 1987
 Persons who have ever undergone long-term hemodialysis
 Persons with unexplained elevations of aminotransferase levels
 Transfusion or transplantation recipients prior to July 1992
 Recipients of blood or organs from a donor found to be positive for hepatitis C
 Children born to women with hepatitis C
 Health care, public safety, and emergency medical personnel following needle injury or mucosal exposure to HCV-contaminated blood
 Sexual partners of persons with hepatitis C infection

Hepatitis E

This type of hepatitis, identified in India, Asia, Africa, the Middle East, and Central America, resembles hepatitis A in its primarily enteric mode of spread. The commonly recognized cases occur after contamination of water supplies such as after monsoon flooding, but sporadic, isolated cases occur. An epidemiologic feature that distinguishes HEV from other enteric agents is the rarity of secondary person-to-person spread from infected persons to their close contacts. Large waterborne outbreaks in endemic areas are linked to genotypes 1 and 2, arise in populations that are immune to HAV, favor young adults, and account for antibody prevalences of 30–80%. In nonendemic areas of the world, such as the United States, clinically apparent acute hepatitis E is extremely rare; however, during the 1988–1994 NHANES survey conducted by the U.S. Public Health Service, the prevalence of anti-HEV was 21%, reflecting subclinical infections, infection with genotypes 3 and 4, predominantly in older males (>60 years). In nonendemic areas, HEV accounts hardly

at all for cases of sporadic (labeled “autochthonous” or indigenous) hepatitis; however, cases imported from endemic areas have been found in the United States. Evidence supports a zoonotic reservoir for HEV primarily in swine, which may account for the mostly subclinical infections in nonendemic areas. A previously unrecognized high distribution of HEV infection, linked to pork-product ingestion, has been discovered in western Europe (e.g., in Germany, an estimated annual incidence of 300,000 cases and a 17% prevalence of anti-HEV among adults; in France, a 22% prevalence of anti-HEV in healthy blood donors).

CLINICAL AND LABORATORY FEATURES

Symptoms and Signs

Acute viral hepatitis occurs after an incubation period that varies according to the responsible agent. Generally, incubation periods for hepatitis A range from 15 to 45 days (mean, 4 weeks), for hepatitis B and D from 30 to 180 days (mean, 8–12 weeks), for hepatitis C from 15 to 160 days (mean, 7 weeks), and for hepatitis E from 14 to 60 days (mean, 5–6 weeks). The *prodromal symptoms* of acute viral hepatitis are systemic and quite variable. Constitutional symptoms of anorexia, nausea and vomiting, fatigue, malaise, arthralgias, myalgias, headache, photophobia, pharyngitis, cough, and coryza may precede the onset of jaundice by 1–2 weeks. The nausea, vomiting, and anorexia are frequently associated with alterations in olfaction and taste. A low-grade fever between 38° and 39°C (100°–102°F) is more often present in hepatitis A and E than in hepatitis B or C, except when hepatitis B is heralded by a serum sickness–like syndrome; rarely, a fever of 39.5°–40°C (103°–104°F) may accompany the constitutional symptoms. Dark urine and clay-colored stools may be noticed by the patient from 1–5 days before the onset of clinical jaundice.

With the onset of *clinical jaundice*, the constitutional prodromal symptoms usually diminish, but in some patients, mild weight loss (2.5–5 kg) is common and may continue during the entire icteric phase. The liver becomes enlarged and tender and may be associated with right upper quadrant pain and discomfort. Infrequently, patients present with a cholestatic picture, suggesting extrahepatic biliary obstruction. Splenomegaly and cervical adenopathy are present in 10–20% of patients with acute hepatitis. Rarely, a few spider angiomas appear during the icteric phase and disappear during convalescence. During the *recovery phase*, constitutional symptoms disappear, but usually some liver enlargement and abnormalities in liver biochemical tests are still evident. The duration of the posticteric phase is variable, ranging from 2 to 12 weeks, and is usually more prolonged in acute hepatitis B and C. Complete clinical and biochemical recovery is to be expected 1–2 months after all cases of hepatitis A and E and 3–4 months after the onset of jaundice in three-quarters of uncomplicated, self-limited cases of hepatitis B and C (among healthy adults, acute hepatitis B is self-limited in 95–99%, whereas hepatitis C is self-limited in only ~15–20%). In the remainder, biochemical recovery may be delayed. A substantial proportion of patients with viral hepatitis never become icteric.

Infection with HDV can occur in the presence of acute or chronic HBV infection; the duration of HBV infection determines the duration of HDV infection. When acute HDV and HBV infections occur simultaneously, clinical and biochemical features may be indistinguishable from those of HBV infection alone, although occasionally they are more severe. As opposed to patients with *acute* HBV infection, patients with *chronic* HBV infection can support HDV replication indefinitely, as when acute HDV infection occurs in the presence of a nonresolving acute HBV infection or, more commonly, when acute hepatitis D is superimposed on underlying chronic hepatitis B. In such cases, the HDV superinfection appears as a clinical exacerbation or an episode resembling acute viral hepatitis in someone already chronically infected with HBV. Superinfection with HDV in a patient with chronic hepatitis B often leads to clinical deterioration (see below).

In addition to superinfections with other hepatitis agents, acute hepatitis-like clinical events in persons with chronic hepatitis B may accompany spontaneous HBeAg to anti-HBe seroconversion or spontaneous reactivation (i.e., reversion from relatively nonreplicative to replicative infection). Such reactivations can occur as well in therapeutically immunosuppressed patients with chronic HBV infection when cytotoxic/immunosuppressive drugs are withdrawn; in these cases, restoration of immune competence is thought to allow resumption of previously checked cell-mediated immune cytolysis of HBV-infected hepatocytes. Occasionally, acute clinical exacerbations of chronic hepatitis B may represent the emergence of a precore mutant (see “[Virology and Etiology](#)”), and the subsequent course in such patients may be characterized by periodic exacerbations. Cytotoxic chemotherapy can lead to reactivation of chronic hepatitis C as well, and anti-TNF- α therapy can lead to reactivation of both hepatitis B and C.

Laboratory Features

The serum aminotransferases aspartate aminotransferase (AST) and ALT (previously designated SGOT and SGPT) increase to a variable degree during the prodromal phase of acute viral hepatitis and precede the rise in bilirubin level ([Figs. 332-2](#) and [332-4](#)). The level of these enzymes, however, does not correlate well with the degree of liver cell damage. Peak levels vary from ~400 to ~4000 IU or more; these levels are usually reached at the time the patient is clinically icteric and diminish progressively during the recovery phase of acute hepatitis. The diagnosis of anicteric hepatitis is based on clinical features and on aminotransferase elevations.

Jaundice is usually visible in the sclera or skin when the serum bilirubin value is $>43 \mu\text{mol/L}$ (2.5 mg/dL). When jaundice appears, the serum bilirubin typically rises to levels ranging from 85 to 340 $\mu\text{mol/L}$ (5–20 mg/dL). The serum bilirubin may continue to rise despite falling serum aminotransferase levels. In most instances, the total bilirubin is equally divided between the conjugated and unconjugated fractions. Bilirubin levels $>340 \mu\text{mol/L}$ (20 mg/dL) extending and persisting late into the course of viral hepatitis are more likely to be associated with severe disease. In certain patients with underlying hemolytic anemia, however, such as glucose-6-phosphate dehydrogenase deficiency and sickle cell anemia, a high serum bilirubin level is common, resulting from superimposed hemolysis. In such patients, bilirubin levels $>513 \mu\text{mol/L}$ (30 mg/dL) have been observed and are not necessarily associated with a poor prognosis.

Neutropenia and lymphopenia are transient and are followed by a relative lymphocytosis. Atypical lymphocytes (varying between 2 and 20%) are common during the acute phase. Measurement of the prothrombin time (PT) is important in patients with acute viral hepatitis, because a prolonged value may reflect a severe hepatic synthetic defect, signify extensive hepatocellular necrosis, and indicate a worse prognosis. Occasionally, a prolonged PT may occur with only mild increases in the serum bilirubin and aminotransferase levels. Prolonged nausea and vomiting, inadequate carbohydrate intake, and poor hepatic glycogen reserves may contribute to hypoglycemia noted occasionally in patients with severe viral hepatitis. Serum alkaline phosphatase may be normal or only mildly elevated, whereas a fall in serum albumin is uncommon in uncomplicated acute viral hepatitis. In some patients, mild and transient steatorrhea has been noted, as well as slight microscopic hematuria and minimal proteinuria.

A diffuse but mild elevation of the γ globulin fraction is common during acute viral hepatitis. Serum IgG and IgM levels are elevated in about one-third of patients during the acute phase of viral hepatitis, but the serum IgM level is elevated more characteristically during acute hepatitis A. During the acute phase of viral hepatitis, antibodies to smooth muscle and other cell constituents may be present, and low titers of rheumatoid factor, nuclear antibody, and heterophile antibody can also be found occasionally. In hepatitis C and D, antibodies to LKM may occur; however, the species of LKM antibodies in the two types of hepatitis are different from each other as well as from the LKM antibody species characteristic of autoimmune hepatitis type 2 ([Chap. 334](#)). The autoantibodies in viral hepatitis are nonspecific and can

also be associated with other viral and systemic diseases. In contrast, virus-specific antibodies, which appear during and after hepatitis virus infection, are serologic markers of diagnostic importance.

As described above, serologic tests are available routinely with which to establish a diagnosis of hepatitis A, B, D, and C. Tests for fecal or serum HAV are not routinely available. Therefore, a diagnosis of hepatitis A is based on detection of IgM anti-HAV during acute illness ([Fig. 332-2](#)). Rheumatoid factor can give rise to false-positive results in this test.

A diagnosis of HBV infection can usually be made by detection of HBsAg in serum. Infrequently, levels of HBsAg are too low to be detected during acute HBV infection, even with contemporary, highly sensitive immunoassays. In such cases, the diagnosis can be established by the presence of IgM anti-HBc.

The titer of HBsAg bears little relation to the severity of clinical disease. Indeed, an inverse correlation exists between the serum concentration of HBsAg and the degree of liver cell damage. For example, titers are highest in immunosuppressed patients, lower in patients with chronic liver disease (but higher in mild chronic than in severe chronic hepatitis), and very low in patients with acute fulminant hepatitis. These observations suggest that in hepatitis B the degree of liver cell damage and the clinical course are related to variations in the patient's immune response to HBV rather than to the amount of circulating HBsAg. In immunocompetent persons, however, a correlation exists between markers of HBV *replication* and liver injury (see below).

Another important serologic marker in patients with hepatitis B is HBeAg. Its principal clinical usefulness is as an indicator of relative infectivity. Because HBeAg is invariably present during early acute hepatitis B, HBeAg testing is indicated primarily in chronic infection.

In patients with hepatitis B surface antigenemia of unknown duration (e.g., blood donors found to be HBsAg-positive) testing for IgM anti-HBc may be useful to distinguish between acute or recent infection (IgM anti-HBc-positive) and chronic HBV infection (IgM anti-HBc-negative, IgG anti-HBc-positive). A false-positive test for IgM anti-HBc may be encountered in patients with high-titer rheumatoid factor. Also, IgM anti-HBc may be reexpressed during acute reactivation of chronic hepatitis B.

Anti-HBs is rarely detectable in the presence of HBsAg in patients with *acute* hepatitis B, but 10–20% of persons with *chronic* HBV infection may harbor low-level anti-HBs. This antibody is directed not against the common group determinant, *a*, but against the heterotypic subtype determinant (e.g., HBsAg of subtype *ad* with anti-HBs of subtype *y*). In most cases, this serologic pattern cannot be attributed to infection with two different HBV subtypes but, instead, is thought (based on the clonal selection theory of antibody diversity) to reflect the stimulation of a related clone of antibody-forming cells and is not a harbinger of imminent HBsAg clearance. When such antibody is detected, its presence is of no recognized clinical significance (see “[Virology and Etiology](#)”).

After immunization with hepatitis B vaccine, which consists of HBsAg alone, anti-HBs is the only serologic marker to appear. The commonly encountered serologic patterns of hepatitis B and their interpretations are summarized in [Table 332-5](#). Tests for the detection of HBV DNA in liver and serum are now available. Like HBeAg, serum HBV DNA is an indicator of HBV replication, but tests for HBV DNA are more sensitive and quantitative. First-generation hybridization assays for HBV DNA had a sensitivity of 10^5 – 10^6 virions/mL, a relative threshold below which infectivity and liver injury are limited and HBeAg is usually undetectable. Currently, testing for HBV DNA has shifted from insensitive hybridization assays to amplification assays (e.g., the PCR-based assay, which can detect as few as 10 or 100 virions/mL); among the commercially available PCR assays, the most useful are those with the highest sensitivity (5–10 IU/mL) and the largest dynamic range (10^0 – 10^9 IU/mL). With increased sensitivity, amplification assays remain reactive well below the current

10^3 – 10^4 IU/mL threshold for infectivity and liver injury. These markers are useful in following the course of HBV replication in patients with chronic hepatitis B receiving antiviral chemotherapy (**Chap. 334**). Except for the early decades of life after perinatally acquired HBV infection (see above), in immunocompetent adults with chronic hepatitis B, a general correlation exists between the level of HBV replication, as reflected by the level of serum HBV DNA, and the degree of liver injury. High-serum HBV DNA levels, increased expression of viral antigens, and necroinflammatory activity in the liver go hand in hand unless immunosuppression interferes with cytolytic T cell responses to virus-infected cells; reduction of HBV replication with antiviral drugs tends to be accompanied by an improvement in liver histology. Among patients with chronic hepatitis B, high levels of HBV DNA increase the risk of cirrhosis, hepatic decompensation, and hepatocellular carcinoma (see “**Complications and Sequelae**”).

TABLE 332-5

Commonly Encountered Serologic Patterns of Hepatitis B Infection

HBsAg	ANTI-HBs	ANTI-HBc	HBeAg	ANTI-HBe	INTERPRETATION
+	–	IgM	+	–	Acute hepatitis B, high infectivity ^a
+	–	IgG	+	–	Chronic hepatitis B, high infectivity
+	–	IgG	–	+	1. Late acute or chronic hepatitis B, low infectivity 2. HBeAg-negative (“precore-mutant”) hepatitis B (chronic or, rarely, acute)
+	+	+	+/-	+/-	1. HBsAg of one subtype and heterotypic anti-HBs (common) 2. Process of seroconversion from HBsAg to anti-HBs (rare)
–	–	IgM	+/-	+/-	1. Acute hepatitis B ^a 2. Anti-HBc “window”
–	–	IgG	–	+/-	1. Low-level hepatitis B carrier 2. Hepatitis B in remote past
–	+	IgG	–	+/-	Recovery from hepatitis B
–	+	–	–	–	1. Immunization with HBsAg (after vaccination) 2. Hepatitis B in the remote past (?) 3. False-positive

^aIgM anti-HBc may reappear during acute reactivation of chronic hepatitis B.

Note: See text for abbreviations.

In patients with hepatitis C, an episodic pattern of aminotransferase elevation is common. A specific serologic diagnosis of hepatitis C can be made by demonstrating the presence in serum of anti-HCV. When contemporary immunoassays are

used, anti-HCV can be detected in acute hepatitis C during the initial phase of elevated aminotransferase activity and remains detectable after recovery (rare) and during chronic infection (common). Nonspecificity can confound immunoassays for anti-HCV, especially in persons with a low prior probability of infection, such as volunteer blood donors, or in persons with circulating rheumatoid factor, which can bind nonspecifically to assay reagents; testing for HCV RNA can be used in such settings to distinguish between true-positive and false-positive anti-HCV determinations. Assays for HCV RNA are the most sensitive tests for HCV infection and represent the “gold standard” in establishing a diagnosis of hepatitis C. HCV RNA can be detected even before acute elevation of aminotransferase activity and before the appearance of anti-HCV in patients with acute hepatitis C. In addition, HCV RNA remains detectable indefinitely, continuously in most but intermittently in some, in patients with chronic hepatitis C (detectable as well in some persons with normal liver tests, i.e., inactive carriers). In the very small minority of patients with hepatitis C who lack anti-HCV, a diagnosis can be supported by detection of HCV RNA. If all these tests are negative and the patient has a well-characterized case of hepatitis after percutaneous exposure to blood or blood products, a diagnosis of hepatitis caused by an unidentified agent can be entertained.

Amplification techniques are required to detect HCV RNA. Currently, such target amplification (i.e., synthesis of multiple copies of the viral genome) is achieved by PCR, in which the viral RNA is reverse transcribed to complementary DNA and then amplified by repeated cycles of DNA synthesis. Quantitative PCR assays provide a measurement of relative “viral load”; current PCR assays have a sensitivity of 10 (lower limit of detection)-25 (lower limit of quantitation) IU/mL and a wide dynamic range ($10-10^7$ IU/mL). Determination of HCV RNA level is not a reliable marker of disease severity or prognosis but is helpful in predicting relative responsiveness to antiviral therapy. The same is true for determinations of HCV genotype (**Chap. 334**). Of course, HCV RNA monitoring during and after antiviral therapy is the *sine qua non* for determining on-treatment and durable responsiveness.

A proportion of patients with hepatitis C have isolated anti-HBc in their blood, a reflection of a common risk in certain populations of exposure to multiple bloodborne hepatitis agents. The anti-HBc in such cases is almost invariably of the IgG class and usually represents HBV infection in the remote past (HBV DNA undetectable); it rarely represents current HBV infection with low-level virus carriage. Detectable anti-HCV in the absence of HCV RNA signifies spontaneous or therapeutically induced recovery from (“cured”) hepatitis C.

The presence of HDV infection can be identified by demonstrating intrahepatic HDV antigen or, more practically, an anti-HDV seroconversion (a rise in titer of anti-HDV or de novo appearance of anti-HDV). Circulating HDV antigen, also diagnostic of acute infection, is detectable only briefly, if at all. Because anti-HDV is often undetectable once HBsAg disappears, retrospective serodiagnosis of acute self-limited, simultaneous HBV and HDV infection is difficult. Early diagnosis of acute infection may be hampered by a delay of up to 30–40 days in the appearance of anti-HDV.

When a patient presents with acute hepatitis and has HBsAg and anti-HDV in serum, determination of the class of anti-HBc is helpful in establishing the relationship between infection with HBV and HDV. Although IgM anti-HBc does not distinguish *absolutely* between acute and chronic HBV infection, its presence is a reliable indicator of recent infection and its absence a reliable indicator of infection in the remote past. In simultaneous acute HBV and HDV infections, IgM anti-HBc will be detectable, whereas in acute HDV infection superimposed on chronic HBV infection, anti-HBc will be of the IgG class. Assays for HDV RNA, available in specialized laboratories and yet to be standardized, can be used to confirm HDV infection and to monitor treatment during chronic infection.

The serologic/virologic course of events during acute hepatitis E is entirely analogous to that of acute hepatitis A, with brief fecal shedding of virus and viremia and an early IgM anti-HEV response that predominates during approximately the first 3 months but is eclipsed thereafter by long-lasting IgG anti-HEV. Diagnostic tests of varying reliability for

hepatitis E are commercially available but used routinely primarily outside the United States; in the United States, diagnostic serologic/virologic assays can be performed at the Centers for Disease Control and Prevention or other specialized reference laboratories.

Liver biopsy is rarely necessary or indicated in acute viral hepatitis, except when the diagnosis is questionable or when clinical evidence suggests a diagnosis of chronic hepatitis.

A diagnostic algorithm can be applied in the evaluation of cases of acute viral hepatitis. A patient with acute hepatitis should undergo four serologic tests: HBsAg, IgM anti-HAV, IgM anti-HBc, and anti-HCV (**Table 332-6**). The presence of HBsAg, with or without IgM anti-HBc, represents HBV infection. If IgM anti-HBc is present, the HBV infection is considered acute; if IgM anti-HBc is absent, the HBV infection is considered chronic. A diagnosis of acute hepatitis B can be made in the absence of HBsAg when IgM anti-HBc is detectable. A diagnosis of acute hepatitis A is based on the presence of IgM anti-HAV. If IgM anti-HAV coexists with HBsAg, a diagnosis of simultaneous HAV and HBV infections can be made; if IgM anti-HBc (with or without HBsAg) is detectable, the patient has simultaneous acute hepatitis A and B, and if IgM anti-HBc is undetectable, the patient has acute hepatitis A superimposed on chronic HBV infection. The presence of anti-HCV supports a diagnosis of acute hepatitis C. Occasionally, testing for HCV RNA or repeat anti-HCV testing later during the illness is necessary to establish the diagnosis. Absence of all serologic markers is consistent with a diagnosis of “non-A, non-B, non-C” hepatitis (no other proven human hepatitis viruses have been identified), if the epidemiologic setting is appropriate.

TABLE 332-6

Simplified Diagnostic Approach in Patients Presenting with Acute Hepatitis

SEROLOGIC TESTS OF PATIENT'S SERUM				
HBsAg	IgM ANTI-HAV	IgM ANTI-HBc	ANTI-HCV	DIAGNOSTIC INTERPRETATION
+	–	+	–	Acute hepatitis B
+	–	–	–	Chronic hepatitis B
+	+	–	–	Acute hepatitis A superimposed on chronic hepatitis B
+	+	+	–	Acute hepatitis A and B
–	+	–	–	Acute hepatitis A
–	+	+	–	Acute hepatitis A and B (HBsAg below detection threshold)
–	–	+	–	Acute hepatitis B (HBsAg below detection threshold)
–	–	–	+	Acute hepatitis C

Note: See text for abbreviations.

In patients with chronic hepatitis, initial testing should consist of HBsAg and anti-HCV. Anti-HCV supports and HCV RNA testing establishes the diagnosis of chronic hepatitis C. If a serologic diagnosis of chronic hepatitis B is made, testing for HBeAg and anti-HBe is indicated to evaluate relative infectivity. Testing for HBV DNA in such patients provides a more quantitative and sensitive measure of the level of virus replication, and therefore is very helpful during antiviral therapy (**Chap. 334**). In patients with chronic hepatitis B and normal aminotransferase activity in the absence of HBeAg, serial testing over time is often required to distinguish between inactive carriage and HBeAg-negative chronic hepatitis B with fluctuating virologic and necroinflammatory activity. In persons with hepatitis B, testing for anti-HDV is useful in those with severe and fulminant disease, with severe chronic disease, with chronic hepatitis B and acute hepatitis-like exacerbations, with frequent percutaneous exposures, and from areas where HDV infection is endemic.

PROGNOSIS

Virtually all previously healthy patients with hepatitis A recover completely with no clinical sequelae. Similarly, in acute hepatitis B, 95–99% of previously healthy adults have a favorable course and recover completely. Certain clinical and laboratory features, however, suggest a more complicated and protracted course. Patients of advanced age and with serious underlying medical disorders may have a prolonged course and are more likely to experience severe hepatitis. Initial presenting features such as ascites, peripheral edema, and symptoms of hepatic encephalopathy suggest a poorer prognosis. In addition, a prolonged PT, low serum albumin level, hypoglycemia, and very high serum bilirubin values suggest severe hepatocellular disease. Patients with these clinical and laboratory features deserve prompt hospital admission. The case fatality rate in hepatitis A and B is very low (~0.1%) but is increased by advanced age and underlying debilitating disorders. Among patients ill enough to be hospitalized for acute hepatitis B, the fatality rate is 1%. Hepatitis C is less severe during the acute phase than hepatitis B and is more likely to be anicteric; fatalities are rare, but the precise case fatality rate is not known. In outbreaks of waterborne hepatitis E in India and Asia, the case fatality rate is 1–2% and up to 10–20% in pregnant women. Contributing to fulminant hepatitis E in endemic countries (but only very rarely or not at all in nonendemic countries) are instances of acute hepatitis E superimposed on underlying chronic liver disease (“acute-on-chronic” liver disease). Patients with simultaneous acute hepatitis B and D do not necessarily experience a higher mortality rate than do patients with acute hepatitis B alone; however, in several outbreaks of acute simultaneous HBV and HDV infection among injection drug users, the case fatality rate was ~5%. When HDV superinfection occurs in a person with chronic hepatitis B, the likelihood of fulminant hepatitis and death is increased substantially. Although the case fatality rate for hepatitis D is not known definitively, in outbreaks of severe HDV superinfection in isolated populations with a high hepatitis B carrier rate, a mortality rate >20% has been recorded.

COMPLICATIONS AND SEQUELAE

A small proportion of patients with hepatitis A experience *relapsing hepatitis* weeks to months after apparent recovery from acute hepatitis. Relapses are characterized by recurrence of symptoms, aminotransferase elevations, occasional jaundice, and fecal excretion of HAV. Another unusual variant of acute hepatitis A is *cholestatic hepatitis*, characterized by protracted cholestatic jaundice and pruritus. Rarely, liver test abnormalities persist for many months, even up to 1 year. Even when these complications occur, hepatitis A remains self-limited and does not progress to chronic liver disease. During the prodromal phase of acute hepatitis B, a serum sickness–like syndrome characterized by arthralgia or arthritis, rash, angioedema, and, rarely, hematuria and proteinuria may develop in 5–10% of patients. This syndrome occurs before the onset of clinical jaundice, and these patients are often diagnosed erroneously as having rheumatologic diseases. The diagnosis can be established by measuring serum aminotransferase levels, which are almost invariably elevated, and serum HBsAg. As noted above, EMC is an immune-complex disease that can complicate chronic hepatitis C and is part of a spectrum of B cell lymphoproliferative disorders, which, in rare instances, can evolve to B cell lymphoma

(**Chap. 104**). Attention has been drawn as well to associations between hepatitis C and such cutaneous disorders as porphyria cutanea tarda and lichen planus. A mechanism for these associations is unknown. Finally, related to the reliance of HCV on lipoprotein secretion and assembly pathways and on interactions of HCV with glucose metabolism, HCV infection may be complicated by hepatic steatosis, hypercholesterolemia, insulin resistance (and other manifestations of the metabolic syndrome), and type 2 diabetes mellitus; both hepatic steatosis and insulin resistance appear to accelerate hepatic fibrosis and blunt responsiveness to interferon-based antiviral therapy (**Chap. 334**).

The most feared complication of viral hepatitis is *fulminant hepatitis* (massive hepatic necrosis); fortunately, this is a rare event. Fulminant hepatitis is seen primarily in hepatitis B, D, and E, but rare fulminant cases of hepatitis A occur primarily in older adults and in persons with underlying chronic liver disease, including, according to some reports, chronic hepatitis B and C. Hepatitis B accounts for >50% of fulminant cases of viral hepatitis, a sizable proportion of which are associated with HDV infection and another proportion with underlying chronic hepatitis C. Fulminant hepatitis is hardly ever seen in hepatitis C, but hepatitis E, as noted above, can be complicated by fatal fulminant hepatitis in 1–2% of all cases and in up to 20% of cases in pregnant women. Patients usually present with signs and symptoms of encephalopathy that may evolve to deep coma. The liver is usually small and the PT excessively prolonged. The combination of rapidly shrinking liver size, rapidly rising bilirubin level, and marked prolongation of the PT, even as aminotransferase levels fall, together with clinical signs of confusion, disorientation, somnolence, ascites, and edema, indicates that the patient has hepatic failure with encephalopathy. Cerebral edema is common; brainstem compression, gastrointestinal bleeding, sepsis, respiratory failure, cardiovascular collapse, and renal failure are terminal events. The mortality rate is exceedingly high (>80% in patients with deep coma), but patients who survive may have a complete biochemical and histologic recovery. If a donor liver can be located in time, liver transplantation may be lifesaving in patients with fulminant hepatitis (**Chap. 338**).

Documenting the disappearance of HBsAg after apparent clinical recovery from acute hepatitis B is particularly important. Before laboratory methods were available to distinguish between acute hepatitis and acute hepatitis-like exacerbations (*spontaneous reactivations*) of chronic hepatitis B, observations suggested that ~10% of previously healthy patients remained HBsAg-positive for >6 months after the onset of clinically apparent acute hepatitis B. One-half of these persons cleared the antigen from their circulations during the next several years, but the other 5% remained chronically HBsAg-positive. More recent observations suggest that the true rate of chronic infection after clinically apparent acute hepatitis B is as low as 1% in normal, immunocompetent, young adults. Earlier, higher estimates may have been confounded by inadvertent inclusion of acute exacerbations in chronically infected patients; these patients, chronically HBsAg-positive before exacerbation, were unlikely to seroconvert to HBsAg-negative thereafter. Whether the rate of chronicity is 10% or 1%, such patients have IgG anti-HBc in serum; anti-HBs is either undetected or detected at low titer against the opposite subtype specificity of the antigen (see “**Laboratory Features**”). These patients may (1) be inactive carriers; (2) have low-grade, mild chronic hepatitis; or (3) have moderate to severe chronic hepatitis with or without cirrhosis. The likelihood of remaining chronically infected after acute HBV infection is especially high among neonates, persons with Down’s syndrome, chronically hemodialyzed patients, and immunosuppressed patients, including persons with HIV infection.

Chronic hepatitis is an important late complication of acute hepatitis B occurring in a small proportion of patients with acute disease but more common in those who present with chronic infection without having experienced an acute illness, as occurs typically after neonatal infection or after infection in an immunosuppressed host (**Chap. 334**). The following clinical and laboratory features suggest progression of acute hepatitis to chronic hepatitis: (1) lack of complete resolution of clinical symptoms of anorexia, weight loss, fatigue, and the persistence of hepatomegaly; (2) the presence of bridging/interface or multilobular hepatic necrosis on liver biopsy during protracted, severe acute viral hepatitis; (3)

failure of the serum aminotransferase, bilirubin, and globulin levels to return to normal within 6–12 months after the acute illness; and (4) the persistence of HBeAg for >3 months or HBsAg for >6 months after acute hepatitis.

Although acute hepatitis D infection does not increase the likelihood of chronicity of simultaneous acute hepatitis B, hepatitis D has the potential for contributing to the severity of chronic hepatitis B. Hepatitis D superinfection can transform inactive or mild chronic hepatitis B into severe, progressive chronic hepatitis and cirrhosis; it also can accelerate the course of chronic hepatitis B. Some HDV superinfections in patients with chronic hepatitis B lead to fulminant hepatitis. As defined in longitudinal studies over three decades, the annual rate of cirrhosis in patients with chronic hepatitis D is 4%. Although HDV and HBV infections are associated with severe liver disease, mild hepatitis and even inactive carriage have been identified in some patients, and the disease may become indolent beyond the early years of infection.

After acute HCV infection, the likelihood of remaining chronically *infected* approaches 85–90%. Although many patients with chronic hepatitis C have no symptoms, cirrhosis may develop in as many as 20% within 10–20 years of acute illness; in some series of cases reported by referral centers, cirrhosis has been reported in as many as 50% of patients with chronic hepatitis C. Among cirrhotic patients with chronic hepatitis C, the annual risk of hepatic decompensation is ~4%. Although chronic hepatitis C accounts for at least 40% of cases of chronic liver disease and of patients undergoing liver transplantation for end-stage liver disease in the United States and Europe, in the majority of patients with chronic hepatitis C, morbidity and mortality are limited during the initial 20 years after the onset of infection. Progression of chronic hepatitis C may be influenced by advanced age of acquisition, long duration of infection, immunosuppression, coexisting excessive alcohol use, concomitant hepatic steatosis, other hepatitis virus infection, or HIV co-infection. In fact, instances of severe and rapidly progressive chronic hepatitis B and C are being recognized with increasing frequency in patients with HIV infection (**Chap. 197**). In contrast, neither HAV nor HEV causes chronic liver disease in immunocompetent hosts; however, cases of chronic hepatitis E (including cirrhosis and end-stage liver disease) have been observed in immunosuppressed organ-transplant recipients, persons receiving cytotoxic chemotherapy, and persons with HIV infection. Among patients with chronic hepatitis (e.g., caused by hepatitis B or C, alcohol, etc.) in endemic countries, hepatitis E has been reported as the cause of acute-on-chronic liver failure; however, in most experiences among patients from nonendemic countries, HEV has not been found to contribute to hepatic decompensation in patients with chronic hepatitis.

Persons with chronic hepatitis B, particularly those infected in infancy or early childhood and especially those with HBeAg and/or high-level HBV DNA, have an enhanced risk of hepatocellular carcinoma. The risks of cirrhosis and hepatocellular carcinoma increase with the level of HBV replication. The annual rate of hepatocellular carcinoma in patients with chronic hepatitis D and cirrhosis is ~3%. The risk of hepatocellular carcinoma is increased as well in patients with chronic hepatitis C, almost exclusively in patients with cirrhosis, and almost always after at least several decades, usually after three decades of disease (**Chap. 78**). Among such cirrhotic patients with chronic hepatitis C, the annual risk of hepatocellular carcinoma is ~1–4%.

Rare complications of viral hepatitis include pancreatitis, myocarditis, atypical pneumonia, aplastic anemia, transverse myelitis, and peripheral neuropathy. In children, hepatitis B may present rarely with anicteric hepatitis, a nonpruritic papular rash of the face, buttocks, and limbs, and lymphadenopathy (papular acrodermatitis of childhood or Gianotti-Crosti syndrome).

Rarely, autoimmune hepatitis (**Chap. 334**) can be triggered by a bout of otherwise self-limited acute hepatitis, as reported after acute hepatitis A, B, and C.

DIFFERENTIAL DIAGNOSIS

Viral diseases such as infectious mononucleosis; those due to cytomegalovirus, herpes simplex, and coxsackieviruses; and toxoplasmosis may share certain clinical features with viral hepatitis and cause elevations in serum aminotransferase and, less commonly, in serum bilirubin levels. Tests such as the differential heterophile and serologic tests for these agents may be helpful in the differential diagnosis if HBsAg, anti-HBc, IgM anti-HAV, and anti-HCV determinations are negative. Aminotransferase elevations can accompany almost any systemic viral infection; other rare causes of liver injury confused with viral hepatitis are infections with *Leptospira*, *Candida*, *Brucella*, *Mycobacteria*, and *Pneumocystis*. A complete drug history is particularly important because many drugs and certain anesthetic agents can produce a picture of either acute hepatitis or cholestasis (**Chap. 333**). Equally important is a past history of unexplained “repeated episodes” of acute hepatitis. This history should alert the physician to the possibility that the underlying disorder is chronic hepatitis, for example autoimmune hepatitis (**Chap. 334**). Alcoholic hepatitis must also be considered, but usually the serum aminotransferase levels are not as markedly elevated, and other stigmata of alcoholism may be present. The finding on liver biopsy of fatty infiltration, a neutrophilic inflammatory reaction, and “alcoholic hyaline” would be consistent with alcohol-induced rather than viral liver injury. Because acute hepatitis may present with right upper quadrant abdominal pain, nausea and vomiting, fever, and icterus, it is often confused with acute cholecystitis, common duct stone, or ascending cholangitis. Patients with acute viral hepatitis may tolerate surgery poorly; therefore, it is important to exclude this diagnosis, and in confusing cases, a percutaneous liver biopsy may be necessary before laparotomy. Viral hepatitis in the elderly is often misdiagnosed as obstructive jaundice resulting from a common duct stone or carcinoma of the pancreas. Because acute hepatitis in the elderly may be quite severe and the operative mortality high, a thorough evaluation including biochemical tests, radiographic studies of the biliary tree, and even liver biopsy may be necessary to exclude primary parenchymal liver disease. Another clinical constellation that may mimic acute hepatitis is right ventricular failure with passive hepatic congestion or hypoperfusion syndromes, such as those associated with shock, severe hypotension, and severe left ventricular failure. Also included in this general category is any disorder that interferes with venous return to the heart, such as right atrial myxoma, constrictive pericarditis, hepatic vein occlusion (Budd-Chiari syndrome), or venoocclusive disease. Clinical features are usually sufficient to distinguish among these vascular disorders and viral hepatitis. Acute fatty liver of pregnancy, cholestasis of pregnancy, eclampsia, and the HELLP (*hemolysis, elevated liver tests, and low platelets*) syndrome can be confused with viral hepatitis during pregnancy. Very rarely, malignancies metastatic to the liver can mimic acute or even fulminant viral hepatitis. Occasionally, genetic or metabolic liver disorders (e.g., Wilson’s disease, α_1 antitrypsin deficiency) and nonalcoholic fatty liver disease are confused with acute viral hepatitis.

TREATMENT

TREATMENT: Acute Viral Hepatitis

Most persons with acute hepatitis (especially hepatitis A, B, and E) recover spontaneously and do not require specific antiviral therapy. In hepatitis B, among previously healthy adults who present with clinically apparent acute hepatitis, recovery occurs in ~99%; therefore, antiviral therapy is not likely to improve the rate of recovery and is not required. In rare instances of severe acute hepatitis B, treatment with a nucleoside analogue at oral doses used to treat chronic hepatitis B (**Chap. 334**) has been attempted successfully. Although clinical trials have not been done to establish the efficacy or duration of this approach, most authorities would recommend institution of antiviral therapy with a nucleoside analogue (entecavir or tenofovir, the most potent and least resistance-prone agents) for severe, but not mild–moderate, acute hepatitis B. Treatment should continue until 3 months after HBsAg seroconversion or 6 months after HBeAg seroconversion.

In typical cases of acute hepatitis C, recovery is rare (~15–20% in most experiences), progression to chronic hepatitis is the rule, and small clinical trials during the era of interferon-based regimens suggested that antiviral therapy with courses (usually 24 weeks) of standard or pegylated interferon α monotherapy reduced the rate of chronicity considerably by inducing sustained responses in 30–70% of patients (according to a meta-analysis of published studies) and in up to 98% in a small German multicenter study (treatment initiated an average of 3 months after infection). In the current interferon-free therapy era, as of 2016, six different all-oral, brief-duration (most lasting 12 weeks), very well-tolerated, highly effective (sustained virologic response rates exceeding 90–95%) combination regimens (of polymerase inhibitors, protease inhibitors, and/or NS5A inhibitors) are available to treat patients with chronic hepatitis C (see Chap. 334); the same regimens are available and recommended to treat patients with acute hepatitis C. Although the duration of therapy for acute hepatitis C has not been determined definitively, in a study of 20 patients, acute hepatitis C resolved after treatment lasting only 6 weeks. In 2016, the European Association for the Study of the Liver (EASL) recommended 8 weeks of treatment for acute hepatitis C with a genotype-appropriate (see Chap. 334) direct-acting antiviral regimen consisting of sofosbuvir plus one of the three approved NS5A inhibitors without ribavirin (12 weeks for patients with acute hepatitis C and either a baseline HCV RNA level >1 million IU/mL or HIV co-infection).

Because spontaneous recovery can occur and because most cases of acute hepatitis C are not clinically severe or rapidly progressive, delaying antiviral therapy of acute hepatitis C for at least 12–16 weeks and even up to 6 months (after which recovery is unlikely) is a recommended approach. Patients with jaundice, those with HCV genotype 1, women, earlier age of infection, lower level of HCV RNA, HBV co-infection, and absence of current injection-drug use are more likely to recover from acute hepatitis C, as are persons who have genetic markers associated with spontaneous recovery (*IL28B* CC haplotype). Because of the marked reduction over the past three decades in the frequency of acute hepatitis C, opportunities to identify and treat patients with acute hepatitis C are rare, except in two population subsets: (1) In health workers who sustain hepatitis C-contaminated needle sticks (occupational accidents), monitoring for ALT elevations and the presence of HCV RNA identify acute hepatitis C in ~3%, and this group should be treated; (2) in injection-drug users, the risk of acute hepatitis C has been on the rise, and the epidemic of opioid use has contributed to an amplification of HCV infection among drug users. Such patients are candidates for antiviral therapy, and efforts to combine antiviral therapy with drug-rehabilitation therapy have been very successful.

Notwithstanding these specific therapeutic considerations, in most cases of typical acute viral hepatitis, specific treatment generally is not necessary. Although hospitalization may be required for clinically severe illness, most patients do not require hospital care. Forced and prolonged bed rest is not essential for full recovery, but many patients will feel better with restricted physical activity. A high-calorie diet is desirable, and because many patients may experience nausea late in the day, the major caloric intake is best tolerated in the morning. Intravenous feeding is necessary in the acute stage if the patient has persistent vomiting and cannot maintain oral intake. Drugs capable of producing adverse reactions such as cholestasis and drugs metabolized by the liver should be avoided. If severe pruritus is present, the use of the bile salt-sequestering resin cholestyramine is helpful. Glucocorticoid therapy has no value in acute viral hepatitis, even in severe cases, and may be deleterious, even increasing the risk of chronicity (e.g., of acute hepatitis B).

Physical isolation of patients with hepatitis to a single room and bathroom is rarely necessary except in the case of fecal incontinence for hepatitis A and E or uncontrolled, voluminous bleeding for hepatitis B (with or without concomitant hepatitis D) and C. Because most patients hospitalized with hepatitis A excrete little, if any, HAV, the likelihood of HAV transmission from these patients during their hospitalization is low. Therefore, burdensome *enteric precautions are no longer recommended*. Although gloves should be worn when the bed pans or fecal material of patients with hepatitis A are handled, these precautions do not represent a departure from sensible procedure and contemporary universal precautions for all hospitalized patients. For patients with hepatitis B and C, emphasis should be placed on blood precautions (i.e., avoiding direct, ungloved hand contact with blood and other body fluids). Enteric precautions are

unnecessary. The importance of simple hygienic precautions such as hand washing cannot be overemphasized. Universal precautions that have been adopted for all patients apply to patients with viral hepatitis. Hospitalized patients may be discharged following substantial symptomatic improvement, a significant downward trend in the serum aminotransferase and bilirubin values, and a return to normal of the PT. Mild aminotransferase elevations should not be considered contraindications to the gradual resumption of normal activity.

In *fulminant hepatitis*, the goal of therapy is to support the patient by maintenance of fluid balance, support of circulation and respiration, control of bleeding, correction of hypoglycemia, and treatment of other complications of the comatose state in anticipation of liver regeneration and repair. Protein intake should be restricted, and oral lactulose administered. Glucocorticoid therapy has been shown in controlled trials to be ineffective. Likewise, exchange transfusion, plasmapheresis, human cross-circulation, porcine liver cross-perfusion, hemoperfusion, and extracorporeal liver-assist devices have not been proven to enhance survival. Meticulous intensive care that includes prophylactic antibiotic coverage is the one factor that appears to improve survival. Orthotopic liver transplantation is resorted to with increasing frequency, with excellent results, in patients with fulminant hepatitis ([Chap. 338](#)). Fulminant hepatitis C is very rare; however, in fulminant hepatitis B, oral antiviral therapy has been used successfully, as reported anecdotally. In clinically severe hepatitis E (with jaundice and coagulopathy), successful therapy with [ribavirin](#) (600 mg twice daily, 15 mg/kg) has been reported anecdotally. Unfortunately, when fulminant hepatitis E occurs in pregnant women (as it does in up to 20% of pregnant women with acute hepatitis E), [ribavirin](#), which is teratogenic, is contraindicated.

PROPHYLAXIS

Because application of therapy for acute viral hepatitis is limited and because antiviral therapy for chronic viral hepatitis is cumbersome, costly, and not effective in all patients ([Chap. 334](#)), emphasis is placed on prevention through immunization. The prophylactic approach differs for each of the types of viral hepatitis. In the past, immunoprophylaxis relied exclusively on passive immunization with antibody-containing globulin preparations purified by cold ethanol fractionation from the plasma of hundreds of normal donors. Currently, for hepatitis A, B, and E, active immunization with vaccines is the preferable approach to prevention.

Hepatitis A

Both passive immunization with IG and active immunization with killed vaccines are available. All preparations of IG contain anti-HAV concentrations sufficient to be protective. When administered before exposure or during the early incubation period, IG is effective in preventing clinically apparent hepatitis A. For postexposure prophylaxis of intimate contacts (household, sexual, institutional) of persons with hepatitis A, the administration of 0.02 mL/kg is recommended as early after exposure as possible; it may be effective even when administered as late as 2 weeks after exposure. Prophylaxis is not necessary for those who have already received hepatitis A vaccine, for casual contacts (office, factory, school, or hospital), for most elderly persons, who are very likely to be immune, or for those known to have anti-HAV in their serum. In day care centers, recognition of hepatitis A in children or staff should provide a stimulus for immunoprophylaxis in the center and in the children's family members. By the time most common-source outbreaks of hepatitis A are recognized, it is usually too late in the incubation period for IG to be effective; however, prophylaxis may limit the frequency of secondary cases. For travelers to tropical countries, developing countries, and other areas outside standard tourist routes, IG prophylaxis had been recommended before a vaccine became available. When such travel lasted <3 months, 0.02 mL/kg was given; for longer travel or residence in these areas, a dose of 0.06 mL/kg every 4–6 months was recommended. Administration of plasma-derived globulin is safe; all contemporary lots of IG are subjected to viral inactivation steps and must be free of HCV RNA as determined by PCR testing. Administration of IM lots of IG has not been associated with transmission of HBV, HCV, or HIV.

Formalin-inactivated vaccines made from strains of HAV attenuated in tissue culture have been shown to be safe, immunogenic, and effective in preventing hepatitis A. Hepatitis A vaccines are approved for use in persons who are at least 1 year old and appear to provide adequate protection beginning 4 weeks after a primary inoculation. If it can be given within 4 weeks of an expected exposure, such as by travel to an endemic area, hepatitis A vaccine is the preferred approach to *preexposure* immunoprophylaxis. If travel is more imminent, IG (0.02 mL/kg) should be administered at a different injection site, along with the first dose of vaccine. Because vaccination provides long-lasting protection (protective levels of anti-HAV should last 20 years after vaccination), persons whose risk will be sustained (e.g., frequent travelers or those remaining in endemic areas for prolonged periods) should be vaccinated, and vaccine should supplant the need for repeated IG injections. Shortly after its introduction, hepatitis A vaccine was recommended for children living in communities with a high incidence of HAV infection; in 1999, this recommendation was extended to include all children living in states, counties, and communities with high rates of HAV infection. As of 2006, the Advisory Committee on Immunization Practices of the U.S. Public Health Service recommended *routine hepatitis A vaccination of all children*. Other groups considered being at increased risk for HAV infection and who are candidates for hepatitis A vaccination include military personnel, populations with cyclic outbreaks of hepatitis A (e.g., Alaskan natives), employees of day care centers, primate handlers, laboratory workers exposed to hepatitis A or fecal specimens, and patients with chronic liver disease. Because of an increased risk of fulminant hepatitis A—observed in some experiences but not confirmed in others—among patients with chronic hepatitis C, patients with chronic hepatitis C are candidates for hepatitis A vaccination, as are persons with chronic hepatitis B. Other populations whose recognized risk of hepatitis A is increased should be vaccinated, including men who have sex with men, injection drug users, persons with clotting disorders who require frequent administration of clotting-factor concentrates, persons traveling from the United States to countries with high or intermediate hepatitis A endemicity, postexposure prophylaxis for contacts of persons with hepatitis A, and household members and other close contacts of adopted children arriving from countries with high and moderate hepatitis A endemicity. Recommendations for dose and frequency differ for the two approved vaccine preparations ([Table 332-7](#)); all injections are IM. Hepatitis A vaccine has been reported to be effective in preventing secondary household and day care center–associated cases of acute hepatitis A. Because the vaccine provides long-lasting protection and is simpler to use, in 2006, the Immunization Practices Advisory Committee of the U.S. Public Health Service favored hepatitis A vaccine to IG for postexposure prophylaxis of healthy persons age 2–40 years; for younger or older persons, for immunosuppressed patients, and for patients with chronic liver disease, IG should continue to be used. In the United States, reported mortality resulting from hepatitis A declined in parallel with hepatitis A vaccine–associated reductions in the annual incidence of new infections.

TABLE 332-7

Hepatitis A Vaccination Schedules

AGE, YEARS	NO. OF DOSES	DOSE	SCHEDULE, MONTHS
HAVRIX (GlaxoSmithKline)^a			
1–18	2	720 ELU ^b (0.5 mL)	0, 6–12
≥19	2	1440 ELU (1 mL)	0, 6–12
VAQTA (Merck)			
1–18	2	25 units (0.5 mL)	0, 6–18
≥19	2	50 units (1 mL)	0, 6–18

^aA combination of this hepatitis A vaccine and hepatitis B vaccine, TWINRIX, is licensed for simultaneous protection against both of these viruses among adults (age ≥18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 µg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6. ^bEnzyme-linked immunoassay units.

Hepatitis B

Until 1982, prevention of hepatitis B was based on *passive* immunoprophylaxis either with standard immunoglobulin, containing modest levels of anti-HBs, or hepatitis B immunoglobulin (HBIG), containing high-titer anti-HBs. The efficacy of standard IG has never been established and remains questionable; even the efficacy of HBIG, demonstrated in several clinical trials, has been challenged, and its contribution appears to be in reducing the frequency of clinical *illness*, not in preventing *infection*. The first vaccine for *active* immunization, introduced in 1982, was prepared from purified, noninfectious 22-nm spherical HBsAg particles derived from the plasma of healthy HBsAg carriers. In 1987, the plasma-derived vaccine was supplanted by a genetically engineered vaccine derived from recombinant yeast. The latter vaccine consists of HBsAg particles that are nonglycosylated but are otherwise indistinguishable from natural HBsAg; two recombinant vaccines are licensed for use in the United States. Current recommendations can be divided into those for preexposure and postexposure prophylaxis.

For *preexposure* prophylaxis against hepatitis B in settings of frequent exposure (health workers exposed to blood; first-responder public safety workers; hemodialysis patients and staff; residents and staff of custodial institutions for the developmentally handicapped; injection drug users; inmates of long-term correctional facilities; persons with multiple sexual partners or who have had a sexually transmitted disease; men who have sex with men; persons such as hemophiliacs who require long-term, high-volume therapy with blood derivatives; household and sexual contacts of persons with chronic HBV infection; persons living in or traveling extensively in endemic areas; unvaccinated children aged <18; unvaccinated children who are Alaskan natives, Pacific Islanders, or residents in households of first-generation immigrants from endemic countries; persons born in countries with a prevalence of HBV infection ≥2%; patients with chronic liver disease; persons < age 60 with diabetes mellitus [those ≥60 at the discretion of their physicians]; persons with end-stage renal disease; and persons with HIV infection), three IM (deltoid, not gluteal) injections of hepatitis B vaccine are recommended at 0, 1, and 6 months (other, optional schedules are summarized in [Table 332-8](#)). Pregnancy is

not a contraindication to vaccination. In areas of low HBV endemicity such as the United States, despite the availability of safe and effective hepatitis B vaccines, a strategy of vaccinating persons in high-risk groups was not effective. The incidence of new hepatitis B cases continued to increase in the United States after the introduction of vaccines; <10% of all targeted persons in high-risk groups were actually vaccinated, and ~30% of persons with sporadic acute hepatitis B did not fall into any high-risk-group category. Therefore, to have an impact on the frequency of HBV infection in an area of low endemicity such as the United States, universal hepatitis B vaccination in childhood has been recommended. For unvaccinated children born after the implementation of universal infant vaccination, vaccination during early adolescence, at age 11–12 years, was recommended, and this recommendation has been extended to include all unvaccinated children age 0–19 years. In HBV-hyperendemic areas (e.g., Asia), universal vaccination of children has resulted in a marked (~70–90%) 30-year decline in complications of hepatitis B, including liver-related mortality and hepatocellular carcinoma.

TABLE 332-8

Preexposure Hepatitis B Vaccination Schedules

TARGET GROUP	NO. OF DOSES	DOSE	SCHEDULE, MONTHS
RECOMBIVAX-HB (Merck)^a			
Infants, children (<1–10 years)	3	5 µg (0.5 mL)	0, 1–2, 4–6
Adolescents (11–19 years)	3 or 4	5 µg (0.5 mL)	0–2, 1–4, 4–6 <i>or</i> 0, 12, 24 <i>or</i> 0, 1, 2, 12
Adults (≥20 years)	<i>Or</i>	10 µg (1 mL)	0, 4–6 (age 11–15)
Hemodialysis patients ^b	2	10 µg (1 mL)	0–2, 1–4, 4–6
<20 years	3	5 µg (0.5 mL)	0, 1, 6
≥20 years	3	40 µg (4 mL)	0, 1, 6
	3		
ENGRIX-B (GlaxoSmithKline)^c			
Infants, children (<1–10 years)	3 or 4	10 µg (0.5 mL)	0, 1–2, 4–6 <i>or</i> 0, 1, 2, 12
Adolescents (10–19 years)	3 or 4	10 µg (0.5 mL)	0, 1–2, 4–6 <i>or</i> 0, 12, 24 <i>or</i> 0, 1, 2, 12
Adults (≥20 years)	3 or 4	20 µg (1 mL)	0–2, 1–4, 4–6 <i>or</i> 0, 1, 2, 12
Hemodialysis patients ^b	4	10 µg (0.5 mL)	0, 1, 2, 6
<20 years	4	40 µg (2 mL)	0, 1, 2, 6
≥20 years			

^aThis manufacturer produces a licensed combination of hepatitis B vaccine and vaccines against *Haemophilus influenzae* type b and *Neisseria meningitidis*, Comvax, for use in infants and young children. Please consult product insert for dose and schedule. ^bThis group also includes other immunocompromised persons. ^cThis manufacturer produces two licensed combination hepatitis B vaccines: (1) Twinrix, recombinant hepatitis B vaccine plus inactivated hepatitis A vaccine, is licensed for simultaneous protection against both of these viruses among adults (age ≥18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 µg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6. (2) Pediarix, recombinant hepatitis B vaccine plus diphtheria and **tetanus toxoid**, pertussis, and inactivated poliovirus, is licensed for use in infants and young children. Please consult product insert for doses and schedules.

The two available recombinant hepatitis B vaccines are comparable, one containing 10 µg of HBsAg (Recombivax-HB) and the other containing 20 µg of HBsAg (Engerix-B), and recommended doses for each injection vary for the two preparations ([Table 332-8](#)). Combinations of hepatitis B vaccine with other childhood vaccines are available as well ([Table 332-8](#)).

For unvaccinated persons sustaining an exposure to HBV, *postexposure* prophylaxis with a combination of HBIG (for rapid achievement of high-titer circulating anti-HBs) and hepatitis B vaccine (for achievement of long-lasting immunity as well as its apparent efficacy in attenuating clinical illness after exposure) is recommended. For *perinatal* exposure of infants born to HBsAg-positive mothers, a single dose of HBIG, 0.5 mL, should be administered IM in the thigh *immediately after birth*, followed by a complete course of three injections of recombinant hepatitis B vaccine (see doses

above) to be started within the first 12 h of life. For those experiencing a direct percutaneous inoculation or transmucosal exposure to HBsAg-positive blood or body fluids (e.g., accidental *needle stick*, other mucosal penetration, or ingestion), a single IM dose of HBIG, 0.06 mL/kg, administered as soon after exposure as possible, is followed by a complete course of hepatitis B vaccine to begin within the first week. For pregnant mothers with high-level HBV DNA ($>2 \times 10^5$ IU/mL), adding antiviral nucleoside analogues (e.g., pregnancy class B tenofovir, see [Chap 334](#)) during the third trimester of pregnancy reduces perinatal transmission even further. For persons exposed by *sexual* contact to a patient with acute hepatitis B, a single IM dose of HBIG, 0.06 mL/kg, should be given within 14 days of exposure, to be followed by a complete course of hepatitis B vaccine. When both HBIG and hepatitis B vaccine are recommended, they may be given at the same time but at separate sites. Testing adults for anti-HBs after a course of vaccine is advisable to document the acquisition of immunity, but, because hepatitis B vaccine immunogenicity is nearly universal in infants, postvaccination anti-HBs testing of children is not recommended.

The precise duration of protection afforded by hepatitis B vaccine is unknown; however, ~80–90% of immunocompetent adult vaccinees retain protective levels of anti-HBs for at least 5 years, and 60–80% for 10 years, and protective antibody has been documented to last for at least two decades after vaccination in infancy. Thereafter and even after anti-HBs becomes undetectable, protection persists against clinical hepatitis B, hepatitis B surface antigenemia, and chronic HBV infection. Currently, *booster* immunizations are not recommended routinely, except in immunosuppressed persons who have lost detectable anti-HBs or immunocompetent persons who sustain percutaneous HBsAg-positive inoculations after losing detectable antibody. Specifically, for hemodialysis patients, annual anti-HBs testing is recommended after vaccination; booster doses are recommended when anti-HBs levels fall to <10 mIU/mL. As noted above, for persons at risk of both hepatitis A and B, a combined vaccine is available containing 720 enzyme-linked immunoassay units (ELUs) of inactivated HAV and 20 μ g of recombinant HBsAg (at 0, 1, and 6 months).

Hepatitis D

Infection with hepatitis D can be prevented by vaccinating susceptible persons with hepatitis B vaccine. No product is available for immunoprophylaxis to prevent HDV superinfection in persons with chronic HBV infection; for them, avoidance of percutaneous exposures and limitation of intimate contact with persons who have HDV infection are recommended.

Hepatitis C

IG is ineffective in preventing hepatitis C and is no longer recommended for postexposure prophylaxis in cases of perinatal, needle stick, or sexual exposure. Although prototype vaccines that induce antibodies to HCV envelope proteins have been developed, currently, hepatitis C vaccination is not feasible practically. Genotype and quasispecies viral heterogeneity, as well as rapid evasion of neutralizing antibodies by this rapidly mutating virus, conspire to render HCV a difficult target for immunoprophylaxis with a vaccine. Prevention of transfusion-associated hepatitis C has been accomplished by the following successively introduced measures: exclusion of commercial blood donors and reliance on a volunteer blood supply; screening donor blood with surrogate markers such as ALT (no longer recommended) and anti-HBc, markers that identify segments of the blood donor population with an increased risk of bloodborne infections; exclusion of blood donors in high-risk groups for AIDS and the introduction of anti-HIV screening tests; and progressively sensitive serologic and virologic screening tests for HCV infection.

In the absence of active or passive immunization, prevention of hepatitis C includes behavior changes and precautions to limit exposures to infected persons. Recommendations designed to identify patients with clinically inapparent hepatitis as candidates for medical management have as a secondary benefit the identification of persons whose

contacts could be at risk of becoming infected. A so-called look-back program has been recommended to identify persons who were transfused before 1992 with blood from a donor found subsequently to have hepatitis C. In addition, anti-HCV testing is recommended for persons born between 1945 and 1965, anyone who received a blood transfusion or a transplanted organ before the introduction of second-generation screening tests in 1992, those who ever used injection drugs (or took other illicit drugs by noninjection routes), chronically hemodialyzed patients, persons with clotting disorders who received clotting factors made before 1987 from pooled blood products, persons with elevated aminotransferase levels, health workers exposed to HCV-positive blood or contaminated needles, recipients of blood or organs from a donor found to be positive for hepatitis C, persons with HIV infection, health care and public safety personnel following a needle stick or other nonpercutaneous exposure to HCV-infected material, sexual partners of persons with hepatitis C, and children born to HCV-positive mothers (Table 332-4).

For stable, monogamous sexual partners, sexual transmission of hepatitis C is unlikely, and sexual barrier precautions are not recommended. For persons with multiple sexual partners or with sexually transmitted diseases, the risk of sexual transmission of hepatitis C is increased, and barrier precautions (latex condoms) are recommended. A person with hepatitis C should avoid sharing such items as razors, toothbrushes, and nail clippers with sexual partners and family members. No special precautions are recommended for babies born to mothers with hepatitis C, and breast-feeding does not have to be restricted.

Hepatitis E

Whether IG prevents hepatitis E remains undetermined. Safe and effective recombinant genotype 1 vaccines, which protect against other genotypes as well, have been developed and are available in endemic areas but not in the United States. Protection provided by the Chinese hepatitis E vaccine is long-lasting, documented in a clinical trial up to 4.5 years.

FURTHER READING

Buckley GJ, Strom BL (eds): *Eliminating the Public Health Problem of Hepatitis B and C in the United States: Phase One Report*. Washington DC, National Academies Press, 2016. Available from <http://www.nap.edu/23407>. Accessed September 22, 2016.

Centers for Disease Control and Prevention: Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945–1965. *MMWR* 61(RR-4):1, 2012.

Debing Y et al: Update on hepatitis E virology: Implications for clinical practice. *J Hepatol* 65:200, 2016. [PubMed: 26966047]

Denniston MM et al: Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003–2010. *Ann Intern Med* 160:293, 2014. [PubMed: 24737271]

Douam F et al: The mechanism of HCV entry into host cells. *Prog Mol Biol Transl Sci* 129:63, 2015. [PubMed: 25595801]

Joy JB et al: The spread of hepatitis C virus genotype 1a in North America: A retrospective phylogenetic study. *Lancet Infect Dis* 16:698, 2016. [PubMed: 27039040]

Lin H-H et al: Changing hepatitis D virus epidemiology in a hepatitis B virus endemic area with a national vaccination program. *Hepatology* 61:1870, 2016.

Pan CQ et al: Tenofovir to prevent hepatitis B transmission in mothers with high viral load. *N Engl J Med* 374:2324, 2016. [[PubMed: 27305192](#)]

Roberts H et al: Prevalence of chronic hepatitis B virus (HBV) infection in U.S. households: National Health and Nutrition Examination Survey (NHANES), 1988–2012. *Hepatology* 63:388, 2016. [[PubMed: 26251317](#)]

Trépo C et al: Hepatitis B virus infection. *Lancet* 384:2053, 2014. [[PubMed: 24954675](#)]

ONLINE BIBLIOGRAPHY

Bartosch B: Hepatitis C virus and its complex interplay with hepatic glucose and lipid metabolism. *J Hepatol* 50:845, 2009. [[PubMed: 19329215](#)]

Centers for Disease Control and Prevention: A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP) part 1: Immunization of infants, children, and adolescents. *MMWR* 54 (RR-16):1, 2005; part II: Immunization of adults. *MMWR* 55(RR-16):1, 2006.

Centers for Disease Control and Prevention: Prevention of hepatitis A through active or passive immunization: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR* 55 (RR-7):1, 2006.

Centers for Disease Control and Prevention: Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. *MMWR* 57(RR-8):1, 2008.

Centers for Disease Control and Prevention: Updated recommendations from the Advisory Committee on Immunization Practices (ACIP) for use of hepatitis A vaccine in close contacts of newly arriving international adoptees. *MMWR* 58:1006, 2009. [[PubMed: 19763077](#)]

Chang M-H et al: Long-term effects of hepatitis B immunization of infants in preventing liver cancer. *Gastroenterology* 151:472, 2016. [[PubMed: 27269245](#)]

Dienstag JL: Hepatitis B virus infection. *N Engl J Med* 359:1486, 2008. [[PubMed: 18832247](#)]

Holmberg SD et al: Hepatitis C in the United States. *N Engl J Med* 368:1859, 2013. [[PubMed: 23675657](#)]

Hughes SA et al: Hepatitis delta virus. *Lancet* 378:73, 2011. [[PubMed: 21511329](#)]

Lindenbach BD, Rice CM: Unravelling hepatitis C virus replication from genome to function. *Nature* 436:933, 2005. [[PubMed: 16107832](#)]

Lindenbach BD et al: Complete replication of hepatitis C virus in cell culture. Science 309:623, 2005. [[PubMed: 15947137](#)]

Ly KN et al: Rising mortality associated with hepatitis C virus in the United States 2003–2013. Clin Infect Dis 62:1287, 2016. [[PubMed: 26936668](#)]

Miller RH et al: Compact organization of the hepatitis B virus genome. Hepatology 9:322, 1989. [[PubMed: 2643549](#)]

Moradpour D et al: Replication of hepatitis C virus. Nature Rev Microbiol 5:453, 2007.

Murphy TV et al: Progress toward eliminating hepatitis A disease in the United States. MMWR 65:29, 2016.

Rehermann B: Pathogenesis of chronic viral hepatitis: differential role of T cells and NK cells. Nature Med 19:859, 2013. [[PubMed: 23836236](#)]

Rizzetto M, Alavian SM: Hepatitis delta: the rediscovery. Clin Liver Dis 17:475, 2013. [[PubMed: 23905817](#)]

Seeger C et al: Biochemical and genetic evidence for the hepatitis B virus replication strategy. Science 232:477, 1986. [[PubMed: 3961490](#)]

Zhang J et al: Long-term efficacy of a hepatitis E vaccine. N Engl J Med 372:914, 2015. [[PubMed: 25738667](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 334: Chronic Hepatitis

Jules L. Dienstag

INTRODUCTION

Chronic hepatitis represents a series of liver disorders of varying causes and severity in which hepatic inflammation and necrosis continue for at least 6 months. Milder forms are nonprogressive or only slowly progressive, while more severe forms may be associated with scarring and architectural reorganization, which, when advanced, lead ultimately to cirrhosis. Several categories of chronic hepatitis have been recognized. These include chronic viral hepatitis, drug-induced chronic hepatitis ([Chap. 333](#)), and autoimmune chronic hepatitis. In many cases, clinical and laboratory features are insufficient to allow assignment into one of these three categories; these “idiopathic” cases are also believed to represent autoimmune chronic hepatitis. Finally, clinical and laboratory features of chronic hepatitis are observed occasionally in patients with such hereditary/metabolic disorders as Wilson’s disease ([copper overload](#)), α_1 antitrypsin deficiency ([Chaps. 337 and 408](#)), and nonalcoholic fatty liver disease ([Chap. 336](#)) and even occasionally in patients with alcoholic liver injury ([Chap. 335](#)). Although all types of chronic hepatitis share certain clinical, laboratory, and histopathologic features, chronic viral and chronic autoimmune hepatitis are sufficiently distinct to merit separate discussions. **For discussion of acute hepatitis, see [Chap. 332](#).**

CLASSIFICATION OF CHRONIC HEPATITIS

Common to all forms of chronic hepatitis are histopathologic distinctions based on localization and extent of liver injury. These vary from the milder forms, previously labeled *chronic persistent hepatitis* and *chronic lobular hepatitis*, to the more severe form, formerly called *chronic active hepatitis*. When first defined, these designations were believed to have prognostic implications, which were not corroborated by subsequent observations. Categorization of chronic hepatitis based primarily on histopathologic features has been replaced by a more informative classification based on a combination of clinical, serologic, and histologic variables. Classification of chronic hepatitis is based on (1) its *cause*; (2) its histologic activity, or *grade*; and (3) its degree of progression based on level of fibrosis, or *stage*. Thus, neither clinical features alone nor histologic features—requiring liver biopsy or noninvasive markers of fibrosis—alone are sufficient to characterize and distinguish among the several categories of chronic hepatitis.

CLASSIFICATION BY CAUSE

Clinical and serologic features allow the establishment of a diagnosis of *chronic viral hepatitis*, caused by hepatitis B, hepatitis B plus D, or hepatitis C; *autoimmune hepatitis*, including several subcategories, I and II, based on serologic distinctions; *drug-associated chronic hepatitis*; and a category of unknown cause, or *cryptogenic chronic hepatitis* (**Table 334-1**). These are addressed in more detail below.

TABLE 334-1

Clinical and Laboratory Features of Chronic Hepatitis

TYPE OF HEPATITIS	DIAGNOSTIC TEST(s)	AUTOANTIBODIES	THERAPY
Chronic hepatitis B	HBsAg, IgG anti-HBc, HBeAg, HBV DNA	Uncommon	IFN- α , PEG IFN- α Oral agents: First-line: entecavir, tenofovir Second-line: lamivudine, adefovir, telbivudine
Chronic hepatitis C	Anti-HCV, HCV RNA	Anti-LKM1 ^a	PEG IFN- α plus ribavirin ^b Direct-acting oral agents: sofosbuvir , ledipasvir, velpatasvir ritonavir-boosted paritaprevir, ombitasvir, dasabavir elbasvir, grazoprevir daclatasvir, simeprevir
Chronic hepatitis D	Anti-HDV, HDV RNA, HBsAg, IgG anti-HBc	Anti-LKM3	IFN- α , PEG IFN- α ^c
Autoimmune hepatitis	ANA ^d (homogeneous), anti-LKM1 (\pm) Hyperglobulinemia	ANA, anti-LKM1 anti-SLA ^e	Prednisone , azathioprine
Drug-associated	—	Uncommon	Withdraw drug
Cryptogenic	All negative	None	Prednisone (?), azathioprine (?)

^aAntibodies to liver-kidney microsomes type 1 (autoimmune hepatitis type II and some cases of hepatitis C). ^bSupplanted in almost all cases by combinations of the direct-acting antiviral agents listed (see www.hcvguidelines.org). ^cEarly clinical trials suggested benefit of IFN- α therapy; PEG IFN- α is as effective, if not more so, and has supplanted standard IFN- α .

^dAntinuclear antibody (autoimmune hepatitis type I). ^eAntibodies to soluble liver antigen (autoimmune hepatitis type III).

Abbreviations: HBc, hepatitis B core; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis D virus; IFN- α , interferon α ; IgG, immunoglobulin G; LKM, liver-kidney microsome; PEG IFN- α , pegylated interferon α ; SLA, soluble liver antigen.

CLASSIFICATION BY GRADE

Grade, a histologic assessment of necroinflammatory activity, is based on examination of the liver biopsy. An assessment of important histologic features includes the degree of *periportal necrosis* and the disruption of the limiting plate of periportal hepatocytes by inflammatory cells (so-called *piecemeal necrosis* or *interface hepatitis*); the degree of confluent necrosis that links or forms bridges between vascular structures—between portal tract and portal tract or even more important bridges between portal tract and central vein—referred to as *bridging necrosis*; the degree of hepatocyte degeneration and focal necrosis within the lobule; and the degree of *portal inflammation*. Several scoring systems that take these histologic features into account have been devised, and the most popular are the histologic activity index (HAI), used commonly in the United States, and the METAVIR score, used in Europe (**Table 334-2**). Based on the presence and degree of these features of histologic activity, chronic hepatitis can be graded as mild, moderate, or severe.

TABLE 334-2

Histologic Grading and Staging of Chronic Hepatitis

	HISTOLOGIC ACTIVITY INDEX (HAI) ^a		METAVIR ^b		
HISTOLOGIC FEATURE	SEVERITY	SCORE	SEVERITY	SCORE	
Necroinflammatory Activity (grade)					
Periportal necrosis, including piecemeal necrosis and/or bridging necrosis (BN)	None	0	None	0	
	Mild	1	Mild	1	
	Mild/moderate	2	Moderate	2	
	Moderate	3	Severe	3	
	Severe	4	Bridging necrosis	Yes No	
Intralobular necrosis Confluent	—None	0	None or mild	0	
	—Focal	1		1	
	—Zone 3 some	2		Moderate	2
	—Zone 3 most	3			
	—Zone 3 + BN few	4	Severe		
	—Zone 3 + BN multiple	5			
	—	6			
	Panacinar/multiacinar				
Focal	—None	0			
	—≤1 focus/10× field	1			
	—2–4 foci/10× field	2			
	—5–10 foci/10× field	3			
	—>10 foci/10× field	4			
Portal Inflammation	None	0	A0–A3 ^c		
	Mild	1			
	Moderate	2			
	Moderate/marked	3			
	Marked	4			
	Total	0–18			
Fibrosis (stage)					

	HISTOLOGIC ACTIVITY INDEX (HAI) ^a		METAVIR ^b	
HISTOLOGIC FEATURE	SEVERITY	SCORE	SEVERITY	SCORE
Necroinflammatory Activity (grade)				
None	Total	0	F0	
		1	F1	
		2	F1	
		3	F2	
		4	F3	
		5	F4	
		6	F4	
Portal fibrosis—some		6	4	
Portal fibrosis—most				
Bridging fibrosis—few				
Bridging fibrosis—many				
Incomplete cirrhosis				
Cirrhosis				

^aIshak K, Baptista A, Bianchi L, et al: Histologic grading and staging of chronic hepatitis. J Hepatol 22:696, 1995. ^bBedossa P, Poynard T, French METAVIR Cooperative Study Group: An algorithm for grading activity in chronic hepatitis C. Hepatology 24:289, 1996. ^cNecroinflammatory grade: A0 = none; A1 = mild; A2 = moderate; A3 = severe.

CLASSIFICATION BY STAGE

The stage of chronic hepatitis, which reflects the level of progression of the disease, is based on the degree of hepatic fibrosis. When fibrosis is so extensive that fibrous septa surround parenchymal nodules and alter the normal architecture of the liver lobule, the histologic lesion is defined as *cirrhosis*. Staging is based on the degree of fibrosis as categorized on a numerical scale 0–6 (HAI) or 0–4 (METAVIR) (Table 334-2). Several noninvasive approaches have been introduced to provide approximations of hepatic histologic stage, including serum biomarkers of fibrosis and imaging determinations of liver elasticity.

CHRONIC VIRAL HEPATITIS

Both the enterically transmitted forms of viral hepatitis, hepatitis A and E, are self-limited and do not cause chronic hepatitis (rare reports notwithstanding in which acute hepatitis A serves as a trigger for the onset of autoimmune hepatitis in genetically susceptible patients or in which hepatitis E (Chap. 332) can cause chronic liver disease in immunosuppressed hosts, for example, after liver transplantation). In contrast, the entire

clinicopathologic spectrum of chronic hepatitis occurs in patients with chronic viral hepatitis B and C as well as in patients with chronic hepatitis D superimposed on chronic hepatitis B.

CHRONIC HEPATITIS B

The likelihood of chronicity after acute hepatitis B varies as a function of age. Infection at birth is associated with clinically silent acute infection but a 90% chance of chronic infection, whereas infection in young adulthood in immunocompetent persons is typically associated with clinically apparent acute hepatitis but a risk of chronicity of only ~1%. Most cases of chronic hepatitis B among adults, however, occur in patients who never had a recognized episode of clinically apparent acute viral hepatitis. The degree of liver injury (grade) in patients with chronic hepatitis B is variable, ranging from none in inactive carriers to mild to moderate to severe. Among adults with chronic hepatitis B, histologic features are of prognostic importance. In one long-term study of patients with chronic hepatitis B, investigators found a 5-year survival rate of 97% for patients with mild chronic hepatitis, 86% for patients with moderate to severe chronic hepatitis, and only 55% for patients with chronic hepatitis and postnecrotic cirrhosis. The 15-year survival in these cohorts was 77%, 66%, and 40%, respectively. On the other hand, more recent observations do not allow us to be so sanguine about the prognosis in patients with mild chronic hepatitis; among such patients followed for 1–13 years, progression to more severe chronic hepatitis and cirrhosis has been observed in more than a quarter of cases.

More important to consider than histology alone in patients with chronic hepatitis B is the degree of hepatitis B virus (HBV) replication. As reviewed in [Chap. 332](#), chronic HBV infection can occur in the presence or absence of serum hepatitis B e antigen (HBeAg), and generally, for both HBeAg-reactive and HBeAg-negative chronic hepatitis B, the level of HBV DNA correlates with the level of liver injury and risk of progression. In *HBeAg-reactive chronic hepatitis B*, two phases have been recognized based on the relative level of HBV replication. The relatively *replicative phase* is characterized by the presence in the serum of HBeAg and HBV DNA levels well in excess of 10^3 – 10^4 IU/mL, sometimes exceeding 10^9 IU/mL; by the presence in the liver of detectable intrahepatocyte nucleocapsid antigens (primarily hepatitis B core antigen [HBcAg]); by high infectivity; and by accompanying liver injury. In contrast, the relatively *nonreplicative phase* is characterized by the absence of the conventional serum marker of HBV replication (HBeAg), the appearance of anti-HBe, levels of HBV DNA below a threshold of $\sim 10^3$ IU/mL, the absence of intrahepatocytic HBcAg, limited infectivity, and minimal liver injury. Patients in the relatively replicative phase tend to have more severe chronic hepatitis, whereas those in the relatively nonreplicative phase tend to have minimal or mild chronic hepatitis or to be inactive hepatitis B carriers. The likelihood in a patient with HBeAg-reactive chronic hepatitis B of converting spontaneously from relatively replicative to nonreplicative infection is ~10% per year. Distinctions in HBV replication and in histologic category, however, do not always coincide. In patients with HBeAg-reactive chronic HBV infection, especially when acquired at birth or in early childhood, as recognized commonly in Asian countries, a dichotomy is common between very high levels of HBV replication during the early decades of life (when the level of apparent host immunologic tolerance of HBV is relatively high) and negligible levels of liver injury. Yet despite the relatively immediate, apparently benign nature of liver disease for many decades in this population, in the middle decades, activation of liver injury emerges as what appears to be the relative tolerance of the host to HBV declines, and these patients with childhood-acquired HBV infection are ultimately

at increased risk later in life of cirrhosis, hepatocellular carcinoma (HCC) (**Chap. 78**), and liver-related death. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis B appears in Chap. 332.**



HBeAg-negative chronic hepatitis B (i.e., chronic HBV infection with active virus replication, readily detectable HBV DNA but without HBeAg [anti-HBe-reactive]), is more common than HBeAg-reactive chronic hepatitis B in Mediterranean and European countries and in Asia (and, correspondingly, in HBV genotypes other than A). Compared to patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B have HBV DNA levels several orders of magnitude lower (no more than 10^5 – 10^6 IU/mL) than those observed in the HBeAg-reactive subset. Most such cases represent precore or core-promoter mutations acquired late in the natural history of the disease (mostly early-life onset; age range 40–55 years, older than that for HBeAg-reactive chronic hepatitis B); these mutations prevent translation of HBeAg from the precore component of the HBV genome (precore mutants) or are characterized by downregulated transcription of precore mRNA (core-promoter mutants; **Chap. 332**). Although their levels of HBV DNA tend to be lower than among patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B can have progressive liver injury (complicated by cirrhosis and HCC) and experience episodic reactivation of liver disease reflected in fluctuating levels of aminotransferase activity (“flares”). The biochemical and histologic activity of HBeAg-negative disease tends to correlate closely with levels of HBV replication, unlike the case mentioned above of Asian patients with HBeAg-reactive chronic hepatitis B during the early decades of their HBV infection. Worth reiterating, the level of HBV replication is the most important risk factor for the ultimate development of cirrhosis and HCC in both HBeAg-reactive (beyond the early decades of “relatively nonreplicative” infection) and HBeAg-negative patients. Although levels of HBV DNA are lower and more readily suppressed by therapy to undetectable levels in HBeAg-negative (compared to HBeAg-reactive) chronic hepatitis B, achieving sustained responses that permit discontinuation of antiviral therapy is less likely in HBeAg-negative patients (see below). Inactive carriers are patients with circulating hepatitis B surface antigen (HBsAg), normal serum aminotransferase levels, undetectable HBeAg, and levels of HBV DNA that are either undetectable or present at a threshold of $\leq 10^3$ IU/mL. This serologic profile occurs not only in inactive carriers but also in patients with HBeAg-negative chronic hepatitis B during periods of relative inactivity; distinguishing between the two requires sequential biochemical and virologic monitoring over many months.

The spectrum of *clinical features* of chronic hepatitis B is broad, ranging from asymptomatic infection to debilitating disease or even end-stage, fatal hepatic failure. As noted above, the onset of the disease tends to be insidious in most patients, with the exception of the very few in whom chronic disease follows failure of resolution of clinically apparent acute hepatitis B. **The clinical and laboratory features associated with progression from acute to chronic hepatitis B are discussed in Chap. 332.**

Fatigue is a common symptom, and persistent or intermittent *jaundice* is a common feature in severe or advanced cases. Intermittent deepening of jaundice and recurrence of malaise and anorexia, as well as worsening fatigue, are reminiscent of acute hepatitis; such exacerbations may occur spontaneously, often coinciding with evidence of virologic reactivation; may lead to progressive liver injury; and, when superimposed on well-established cirrhosis, may cause hepatic decompensation. Complications of cirrhosis occur in end-stage chronic hepatitis and include ascites, edema, bleeding gastroesophageal varices, hepatic

encephalopathy, coagulopathy, and hypersplenism. Occasionally, these complications bring the patient to initial clinical attention. Extrahepatic complications of chronic hepatitis B, similar to those seen during the prodromal phase of acute hepatitis B, are associated with tissue deposition of circulating hepatitis B antigen–antibody immune complexes. These include arthralgias and arthritis, which are common, and the more rare purpuric cutaneous lesions (leukocytoclastic vasculitis), immune-complex glomerulonephritis, and generalized vasculitis (polyarteritis nodosa) (**Chaps. 332 and 356**).

Laboratory features of chronic hepatitis B do not distinguish adequately between histologically mild and severe hepatitis. Aminotransferase elevations tend to be modest for chronic hepatitis B but may fluctuate in the range of 100–1000 units. As is true for acute viral hepatitis B, alanine aminotransferase (ALT) tends to be more elevated than aspartate aminotransferase (AST); however, once cirrhosis is established, AST tends to exceed ALT. Levels of alkaline phosphatase activity tend to be normal or only marginally elevated. In severe cases, moderate elevations in serum bilirubin (51.3–171 $\mu\text{mol/L}$ [3–10 mg/dL]) occur. Hypoalbuminemia and prolongation of the prothrombin time occur in severe or end-stage cases. Hyperglobulinemia and detectable circulating autoantibodies are distinctly absent in chronic hepatitis B (in contrast to autoimmune hepatitis).

Viral markers of chronic HBV infection are discussed in Chap. 332.

TREATMENT

TREATMENT: Chronic Hepatitis B

Although progression to cirrhosis is more likely in severe than in mild or moderate chronic hepatitis B, all forms of chronic hepatitis B can be progressive, and progression occurs primarily in patients with active HBV replication. Moreover, in populations of patients with chronic hepatitis B who are at risk for HCC (**Chap. 78**), the risk is highest for those with continued, high-level HBV replication and lower for persons in whom initially high-level HBV DNA falls spontaneously over time. Therefore, management of chronic hepatitis B is directed at suppressing the level of virus replication. Although clinical trials tend to focus on clinical endpoints achieved over 1–2 years (e.g., suppression of HBV DNA to undetectable levels, loss of HBeAg/HBsAg, improvement in histology, normalization of ALT), these short-term gains translate into reductions in the risk of clinical progression, hepatic decompensation, HCC, liver transplantation, and death; regression of cirrhosis and of esophageal varices have been documented to follow long-term pharmacologic suppression of HBV replication. In addition, restoration of impaired HBV-specific T-cell function has been shown following successful suppression of HBV replication with antiviral therapy. To date, seven drugs have been approved for treatment of chronic hepatitis B: injectable interferon (IFN) α and pegylated interferon (long-acting IFN bound to polyethylene glycol, PEG [PEG IFN]) and the oral agents lamivudine, adefovir dipivoxil, entecavir, telbivudine, and tenofovir disoproxil fumarate (TDF).

Antiviral therapy for hepatitis B has evolved rapidly since the mid-1990s, as has the sensitivity of tests for HBV DNA. When IFN and lamivudine were evaluated in clinical trials, HBV DNA was measured by insensitive hybridization assays with detection thresholds of 10^5 – 10^6 virions/mL; when adefovir, entecavir, telbivudine, tenofovir, and PEG IFN were studied in clinical trials, HBV DNA was measured by sensitive amplification assays (polymerase chain reaction [PCR]) with detection thresholds of 10^1 – 10^3 viral copies/mL or IU/mL. Recognition

of these distinctions is helpful when comparing results of clinical trials that established the efficacy of these therapies (reviewed below in chronological order of publication of these efficacy trials).

INTERFERON

IFN- α was the first approved therapy (1992) for chronic hepatitis B. Although it is no longer used to treat hepatitis B, standard IFN is important historically, having provided important lessons about antiviral therapy in general. For immunocompetent adults with HBeAg-reactive chronic hepatitis B (who tend to have high-level HBV DNA [$>10^5$ – 10^6 virions/mL] and histologic evidence of chronic hepatitis on liver biopsy), a 16-week course of IFN given subcutaneously at a daily dose of 5 million units, or three times a week at a dose of 10 million units, resulted in a loss of HBeAg and hybridization-detectable HBV DNA (i.e., a reduction to levels below 10^5 – 10^6 virions/mL) in ~30% of patients, with a concomitant improvement in liver histology. Seroconversion from HBeAg to anti-HBe occurred in ~20%, and, in early trials, ~8% lost HBsAg. Successful IFN therapy and seroconversion were often accompanied by an acute hepatitis-like elevation in aminotransferase activity, postulated to result from enhanced cytolytic T cell clearance of HBV-infected hepatocytes. Relapse after successful therapy was rare (1 or 2%). The likelihood of responding to IFN was higher in patients with lower levels of HBV DNA and substantial elevations of ALT. Although children can respond as well as adults, IFN therapy was not effective in very young children infected at birth. Similarly, IFN therapy was not effective in immunosuppressed persons, Asian patients with neonatal acquisition of infection and minimal-to-mild ALT elevations, or patients with decompensated chronic hepatitis B (in whom such therapy was actually detrimental, sometimes precipitating decompensation, often associated with severe adverse effects). Among patients with HBeAg loss during therapy, long-term follow-up demonstrated that 80% experienced eventual loss of HBsAg (i.e., all serologic markers of infection, and normalization of ALT over a 9-year posttreatment period). In addition, improved long-term and complication-free survival as well as a reduction in the frequency of HCC were documented among IFN responders, supporting the conclusion that successful antiviral therapy improves the natural history of chronic hepatitis B.

Initial trials of brief-duration IFN therapy in patients with *HBeAg-negative chronic hepatitis B* were disappointing, suppressing HBV replication transiently during therapy but almost never resulting in sustained antiviral responses. In subsequent IFN trials among patients with HBeAg-negative chronic hepatitis B, however, more protracted courses, lasting up to 1.5 years, were reported to result in sustained remissions documented to last for several years, with suppressed HBV DNA and aminotransferase activity, in ~20%.

Complications of IFN therapy include systemic “flu-like” symptoms; marrow suppression; emotional lability (irritability, depression, anxiety); autoimmune reactions (especially autoimmune thyroiditis); and miscellaneous side effects such as alopecia, rashes, diarrhea, and numbness and tingling of the extremities. With the possible exception of autoimmune thyroiditis, all these side effects are reversible upon dose lowering or cessation of therapy.

Although no longer competitive with the newer generation of antivirals, IFN did represent the first successful antiviral approach and set a standard against which to measure subsequent drugs in the achievement of durable virologic, serologic, biochemical, and histologic responses; consolidation of virologic and biochemical benefit in the ensuing years after therapy; and improvement in the natural history of chronic hepatitis B.

Standard IFN has been supplanted by long-acting PEG IFN (see below), and IFN nonresponders are now treated with one of the newer oral nucleoside analogues.

LAMIVUDINE

The first of the nucleoside analogues to be approved (in 1998) for hepatitis B, the dideoxynucleoside lamivudine inhibits reverse transcriptase activity of both HIV and HBV and is an effective agent for patients with chronic hepatitis B. Although generally superseded by newer, more potent, less resistance-prone agents, lamivudine is still used in regions of the world where newer agents are not yet available or affordable. In clinical trials among patients with HBeAg-reactive chronic hepatitis B, lamivudine therapy at daily doses of 100 mg for 48–52 weeks suppressed HBV DNA by a median of $\sim 5.5 \log_{10}$ copies/mL and to undetectable levels, as measured by PCR amplification assays, in $\sim 40\%$ of patients. Therapy was associated with HBeAg loss in 32–33%, HBeAg seroconversion (i.e., conversion from HBeAg-reactive to anti-HBe-reactive) in 16–21%, normalization of ALT in 40–75%, improvement in histology in 50–60%, retardation in hepatic fibrosis in 20–30%, and prevention of progression to cirrhosis. HBeAg responses occur even in patients resistant to IFN (e.g., those with high-level HBV DNA) or who failed in the past to respond to it. As is true for IFN therapy of chronic hepatitis B, patients with near-normal ALT activity tend not to experience HBeAg responses (despite suppression of HBV DNA), whereas those with ALT levels exceeding $5 \times$ the upper limit of normal can expect 1-year HBeAg seroconversion rates of 50–60%. Generally, HBeAg seroconversions are confined to patients who achieve suppression of HBV DNA to $<10^4$ copies/mL (equivalent to $\sim 10^3$ IU/mL). Lamivudine-associated HBeAg responses are accompanied by a delayed posttreatment HBsAg seroconversion rate comparable to that seen after IFN-induced HBeAg responses. Among Western patients who undergo HBeAg responses during a year-long course of therapy and in whom the response is sustained for 4–6 months after cessation of therapy, the response is durable thereafter in the vast majority ($>80\%$); therefore, the achievement of an HBeAg response represents a viable stopping point in therapy. Reduced durability has been reported in Asian patients; therefore, to support the durability of HBeAg responses, patients should receive a period of consolidation therapy of ≥ 6 months in Western patients and ≥ 1 year in Asian patients after HBeAg seroconversion (see treatment guidelines below). Close posttreatment monitoring is necessary to identify HBV reactivation promptly and to resume therapy. If HBeAg is unaffected by lamivudine therapy, the current approach is to continue therapy until an HBeAg response occurs, but long-term therapy may be required to suppress HBV replication and, in turn, limit liver injury; HBeAg seroconversions can increase to a level of 50% after 5 years of therapy. Histologic improvement continues to accrue with therapy beyond the first year; after a cumulative course of 3 years of lamivudine therapy, necroinflammatory activity is reduced in the majority of patients, and even cirrhosis has been shown to regress to precirrhotic stages in as many as three-quarters of patients.

Losses of HBsAg have been few during the first year of lamivudine therapy, and this observation had been cited as an advantage of IFN-based over lamivudine therapy; however, in head-to-head comparisons between standard IFN and lamivudine monotherapy, HBsAg losses were rare in both groups. Trials in which lamivudine and IFN were administered in combination failed to show a benefit of combination therapy over lamivudine monotherapy for either treatment-naïve patients or prior IFN nonresponders.

In patients with *HBeAg-negative chronic hepatitis B* (i.e., in those with precore and core-promoter HBV mutations), 1 year of lamivudine therapy results in HBV DNA suppression and normalization of ALT in three-

quarters of patients and in histologic improvement in approximately two-thirds. Therapy has been shown to suppress HBV DNA by $\sim 4.5 \log_{10}$ copies/mL (baseline HBV DNA levels are lower than in patients with HBeAg-reactive hepatitis B) and to undetectable levels in $\sim 70\%$, as measured by sensitive PCR amplification assays. Lacking HBeAg at the outset, patients with HBeAg-negative chronic hepatitis B cannot achieve an HBeAg response—a stopping point in HBeAg-reactive patients; almost invariably, when therapy is discontinued, reactivation is the rule. Therefore, these patients require long-term therapy; with successive years, the proportion with suppressed HBV DNA and normal ALT increases.

Clinical and laboratory side effects of lamivudine are negligible and indistinguishable from those observed in placebo recipients. Still, lamivudine doses should be reduced in patients with reduced creatinine clearance. During lamivudine therapy, transient ALT elevations, resembling those seen during IFN therapy and during spontaneous HBeAg-to-anti-HBe seroconversions, occur in one-fourth of patients. These ALT elevations may result from restored cytolytic T cell activation permitted by suppression of HBV replication. Similar ALT elevations, however, occurred at an identical frequency in placebo recipients; however, ALT elevations associated with HBeAg seroconversion in clinical trials were confined to lamivudine-treated patients. When therapy is stopped after a year of therapy, two- to threefold ALT elevations occur in 20–30% of lamivudine-treated patients, representing renewed liver-cell injury as HBV replication returns. Although these posttreatment flares are almost always transient and mild, rare severe exacerbations, especially in cirrhotic patients, have been observed, mandating close and careful clinical and virologic monitoring after discontinuation of treatment. Many authorities caution against discontinuing therapy in patients with cirrhosis, in whom posttreatment flares could precipitate decompensation.

Long-term monotherapy with lamivudine is associated with methionine-to-valine (M204V) or methionine-to-isoleucine (M204I) mutations, primarily at amino acid 204 in the tyrosine-methionine-aspartate-aspartate (YMDD) motif of the C domain of HBV DNA polymerase, analogous to mutations that occur in HIV-infected patients treated with this drug. During a year of therapy, YMDD mutations occur in 15–30% of patients; the frequency increases with each year of therapy, reaching 70% at year 5. Ultimately, patients with YMDD mutants experience degradation of clinical, biochemical, and histologic responses; therefore, if treatment is begun with lamivudine monotherapy, the emergence of lamivudine resistance, reflected clinically by a breakthrough from suppressed levels of HBV DNA and ALT, is managed by adding another antiviral to which YMDD variants are sensitive (e.g., adefovir, tenofovir; see below).

Currently, although lamivudine is very safe and still used widely in other parts of the world, in the United States and Europe, lamivudine has been eclipsed by more potent antivirals that have superior resistance profiles (see below); it is no longer recommended as first-line therapy. Still, as the first successful oral antiviral agent for use in hepatitis B, lamivudine provided proof of principle that polymerase inhibitors can achieve virologic, serologic, biochemical, and histologic benefits. In addition, lamivudine has been shown to be effective in the treatment of patients with decompensated hepatitis B (for whom IFN is contraindicated), in some of whom decompensation can be reversed. Moreover, among patients with cirrhosis or advanced fibrosis, lamivudine has been shown to be effective in reducing the risk of progression to hepatic decompensation and, marginally, the risk of HCC. In the half decade following the introduction in the United States of lamivudine therapy for hepatitis B, referral of patients with HBV-associated end-stage liver disease for liver transplantation was

reduced by ~30%, supporting further the beneficial impact of oral antiviral therapy on the natural history of chronic hepatitis B.

Because lamivudine monotherapy can result universally in the rapid emergence of YMDD variants in persons with HIV infection, patients with chronic hepatitis B should be tested for anti-HIV prior to therapy; if HIV infection is identified, lamivudine monotherapy at the HBV daily dose of 100 mg is contraindicated. These patients should be treated for both HIV and HBV with an HIV drug regimen that includes or is supplemented by at least two drugs active against HBV; antiretroviral therapy (ART) often contains two drugs with antiviral activity against HBV (e.g., tenofovir and emtricitabine), but if lamivudine is part of the regimen, the daily dose should be 300 mg (**Chap. 197**). The safety of lamivudine during pregnancy has not been established; however, the drug is not teratogenic in rodents and has been used safely in pregnant women with HIV infection and with HBV infection. Administration of lamivudine during the last months of pregnancy to mothers with high-level hepatitis B viremia ($\geq 10^8$ IU/mL) can reduce the likelihood of perinatal transmission of hepatitis B.

ADEFOVIR DIPIVOXIL

At an oral daily dose of 10 mg, the acyclic nucleotide analogue adefovir dipivoxil, the prodrug of adefovir (approved for hepatitis B in 2002), reduces HBV DNA by $\sim 3.5\text{--}4 \log_{10}$ copies/mL and is equally effective in treatment-naïve patients and prior IFN nonresponders. In HBeAg-reactive chronic hepatitis B, a 48-week course of adefovir dipivoxil was shown to achieve histologic improvement (and reduce the progression of fibrosis) and normalization of ALT in just over one-half of patients, HBeAg seroconversion in 12%, HBeAg loss in 23%, and suppression to an undetectable level of HBV DNA in 13–21%, as measured by PCR. Similar to IFN and lamivudine, adefovir dipivoxil is more likely to achieve an HBeAg response in patients with high baseline ALT; among adefovir-treated patients with ALT level $> 5 \times$ the upper limit of normal, HBeAg seroconversions occurred in 25%. The durability of adefovir-induced HBeAg responses is high (91% in one study); therefore, HBeAg response can be relied upon as a stopping point for adefovir therapy, after a period of consolidation therapy, as outlined above. Although data on the impact of additional therapy beyond 1 year are limited, biochemical, serologic, and virologic outcomes improve progressively as therapy is continued.

In patients with *HBeAg-negative chronic hepatitis B*, a 48-week course of 10 mg/d of adefovir dipivoxil resulted in histologic improvement in two-thirds, normalization of ALT in three-fourths, and suppression of HBV DNA to PCR-undetectable levels in one-half to two-thirds. As was true for lamivudine, because HBeAg responses—a potential stopping point—cannot be achieved in this group, reactivation is the rule when adefovir therapy is discontinued, and indefinite, long-term therapy is required. Treatment beyond the first year consolidates the gain of the first year; after 5 years of therapy, improvement in hepatic inflammation and regression of fibrosis were observed in three-fourths of patients, ALT was normal in 70%, and HBV DNA was undetectable in almost 70%. In one study, stopping adefovir after 5 years was followed by sustained suppression of HBV DNA and ALT, but most HBeAg-negative patients are treated indefinitely unless HBsAg loss, albeit very rare, is achieved.

Adefovir contains a flexible acyclic linker instead of the L-nucleoside ring of lamivudine, avoiding steric hindrance by mutated amino acids. In addition, the molecular structure of phosphorylated adefovir is very similar to that of its natural substrate; therefore, mutations to adefovir would also affect binding of the natural substrate, dATP. Hypothetically, these are among the reasons that resistance to adefovir dipivoxil is much less

likely than resistance to lamivudine; no resistance was encountered in 1 year of clinical trial therapy. In subsequent years, however, adefovir resistance begins to emerge (asparagine to threonine at amino acid 236 [N236T] and alanine to valine or threonine at amino acid 181 [A181V/T], primarily), occurring in 2.5% after 2 years, but in 29% after 5 years of therapy (reported in HBeAg-negative patients). Among patients co-infected with HBV and HIV and who have normal CD4+ T cell counts, adefovir dipivoxil is effective in suppressing HBV dramatically (by 5 logs₁₀ in one study). Moreover, adefovir dipivoxil is effective in lamivudine-resistant, YMDD-mutant HBV and can be used when such lamivudine-induced variants emerge. When lamivudine resistance occurs, adding adefovir (i.e., maintaining lamivudine to preempt the emergence of adefovir resistance) is superior to switching to adefovir. Almost invariably, patients with adefovir-induced HBV mutations respond to lamivudine (or newer agents, such as entecavir, see below). When, in the past, adefovir had been evaluated as therapy for HIV infection, doses of 60–120 mg were required to suppress HIV, and, at these doses, the drug was nephrotoxic. Even at 30 mg/d, creatinine elevations of 44 µmol/L (0.5 mg/dL) occurred in 10% of patients; however, at the HBV-effective dose of 10 mg, such creatinine elevations are rarely encountered. If any nephrotoxicity does occur, it rarely appears before 6–8 months of therapy. Although renal tubular injury is a rare potential side effect, and although creatinine monitoring is recommended during treatment, the therapeutic index of adefovir dipivoxil is high, and the nephrotoxicity observed in clinical trials at higher doses was reversible. For patients with underlying renal disease, frequency of administration of adefovir dipivoxil should be reduced to every 48 h for creatinine clearances of 30–49 mL/min; to every 72 h for creatinine clearances of 10–29 mL/min; and to once a week, following dialysis, for patients undergoing hemodialysis. Adefovir dipivoxil is very well tolerated, and ALT elevations during and after withdrawal of therapy are similar to those observed and described above in clinical trials of lamivudine. An advantage of adefovir is its relatively favorable resistance profile; however, it is not as potent as the other approved oral agents, it does not suppress HBV DNA as rapidly or as uniformly as the others, it is the least likely of all agents to result in HBeAg seroconversion, and 20–50% of patients fail to suppress HBV DNA by 2 log₁₀ (“primary nonresponders”). For these reasons, adefovir, which has been supplanted in both treatment-naïve and lamivudine-resistant patients by the more potent, less resistance-prone nucleotide analogue tenofovir (see below), is no longer recommended as first-line therapy.

PEGYLATED IFN

After long-acting PEG IFN was shown to be effective in the treatment of hepatitis C (see below), this more convenient drug was evaluated in the treatment of chronic hepatitis B. Once-a-week PEG IFN is more effective than the more frequently administered, standard IFN, and several large-scale trials of PEG IFN versus oral nucleoside analogues were conducted among patients with HBeAg-reactive and HBeAg-negative chronic hepatitis B.

In HBeAg-reactive chronic hepatitis B, two large-scale studies were done. In one study, PEG IFN-α 2b (100 µg weekly for 32 weeks, then 50 µg weekly for another 20 weeks for a total of 52 weeks) was evaluated against a comparison arm of combination PEG IFN with oral lamivudine in 307 subjects. The other study involved PEG IFN-α 2a (180 µg weekly for 48 weeks) in 814 primarily Asian patients, three-fourths of whom had ALT ≥ 2 × the upper limit of normal, with comparison arms of lamivudine monotherapy and combination PEG IFN plus lamivudine. At the end of therapy (48–52 weeks) in the PEG IFN monotherapy arms, HBeAg loss occurred in

~30%, HBeAg seroconversion in 22–27%, undetectable HBV DNA (<400 copies/mL by PCR) in 10–25%, normal ALT in 34–39%, and a mean reduction in HBV DNA of 2 log₁₀ copies/mL (PEG IFN-α 2b) to 4.5 log₁₀ copies/mL (PEG IFN-α 2a). Six months after completing PEG IFN monotherapy in these trials, HBeAg losses were present in ~35%, HBeAg seroconversion in ~30%, undetectable HBV DNA in 7–14%, normal ALT in 32–41%, and a mean reduction in HBV DNA of 2–2.4 log₁₀ copies/mL. Although the combination of PEG IFN and lamivudine was superior at the end of therapy in one or more serologic, virologic, or biochemical outcomes, neither the combination arm (in both studies) nor the lamivudine monotherapy arm (in the PEG IFN-α 2a trial) demonstrated any benefit compared to the PEG IFN monotherapy arms 6 months after therapy. Moreover, HBsAg seroconversion occurred in 3–7% of PEG IFN recipients (with or without lamivudine); some of these seroconversions were identified by the end of therapy, but many were identified during the posttreatment follow-up period. The likelihood of HBeAg loss in PEG IFN-treated HBeAg-reactive patients is associated with HBV genotype A > B > C > D (shown for PEG IFN-α2b but not for α-2a). PEG IFN-α 2a was approved in the US for hepatitis B in 2005; PEG IFN-α 2b, not approved for hepatitis B in the US, is used in other countries.

Based on these results, some authorities concluded that PEG IFN monotherapy should be the first-line therapy of choice in HBeAg-reactive chronic hepatitis B; however, this conclusion has been challenged. Although a finite, 1-year course of PEG IFN results in a higher rate of sustained response (6 months after treatment) than is achieved with oral nucleoside/nucleotide analogue therapy, the comparison is confounded by the fact that oral agents are not discontinued at the end of 1 year. Instead, taken orally and free of side effects, therapy with oral agents is extended indefinitely or until after the occurrence of an HBeAg response. The rate of HBeAg responses after 2 years of oral-agent nucleoside analogue therapy is at least as high as, if not higher than, that achieved with PEG IFN after 1 year; favoring oral agents is the absence of injections, difficult-to-tolerate side effects, and laboratory monitoring as well as lower direct and indirect medical care costs and inconvenience. The association of HBsAg responses with PEG IFN therapy occurs in such a small proportion of patients that subjecting everyone to PEG IFN for the marginal gain of HBsAg responses during or immediately after therapy in such a very small minority is questionable. Moreover, HBsAg responses occur in a comparable proportion of patients treated with early-generation nucleoside/nucleotide analogues in the years *after* therapy, and, with the newer, more potent nucleoside analogues, the frequency of HBsAg loss during the first year of therapy equals that of PEG IFN and is exceeded during year 2 and beyond (see below). Of course, resistance is not an issue during PEG IFN therapy, but the risk of resistance is much lower with new agents (≤1% up to 3–8 years in previously treatment-naïve, entecavir-treated and 0% of tenofovir-treated patients; see below). Finally, the level of HBV DNA inhibition that can be achieved with the newer agents, and even with lamivudine, exceeds that which can be achieved with PEG IFN, in some cases by several orders of magnitude.

In HBeAg-negative chronic hepatitis B, a trial of PEG IFN-α 2a (180 µg weekly for 48 weeks versus comparison arms of lamivudine monotherapy and of combination therapy) in 564 patients showed that PEG IFN monotherapy resulted at the end of therapy in suppression of HBV DNA by a mean of 4.1 log₁₀ copies/mL, undetectable HBV DNA (<400 copies/mL by PCR) in 63%, normal ALT in 38%, and loss of HBsAg in 4%. Although lamivudine monotherapy and combination lamivudine–PEG IFN therapy were both superior to PEG IFN at the end of therapy, no advantage of lamivudine monotherapy or combination therapy was apparent over PEG IFN monotherapy 6 months after therapy—suppression of HBV DNA by a mean of 2.3 log₁₀ copies/mL, undetectable

HBV DNA in 19%, and normal ALT in 59%. In subjects involved in this trial followed for up to 5 years, among the two-thirds followed who had been treated initially with PEG IFN, 17% maintained HBV DNA suppression to <400 copies/mL, but ALT remained normal in only 22%; HBsAg loss increased gradually to 12%. Among the half followed who had been treated initially with lamivudine monotherapy, HBV DNA remained <400 copies/mL in 7% and ALT normal in 16%; by year 5, 3.5% had lost HBsAg. As was the case for standard IFN therapy in HBeAg-negative patients, only a small proportion maintained responsiveness after completion of PEG IFN therapy, raising questions about the relative value of a finite period of PEG IFN, versus a longer course with a potent, low-resistance oral nucleoside analogue in these patients. Moreover, the value of PEG IFN for HBeAg-negative chronic hepatitis B has not been confirmed. In the only other controlled clinical trial of PEG IFN for HBeAg-negative chronic hepatitis B, the hepatitis C regimen of PEG IFN plus [ribavirin](#) was compared to PEG IFN monotherapy. In this trial, HBV DNA suppression (<400 copies/mL) occurred in only 7.5% of the two groups combined, and no study subject lost HBsAg.

In patients treated with PEG IFN, HBeAg and HBsAg responses have been associated with *IL28B* genotype CC, the favorable genotype identified in trials of PEG IFN for chronic hepatitis C. Also, reductions in quantitative HBsAg levels have been shown to correlate with and to be predictive of responsiveness to PEG IFN in chronic hepatitis B. If HBsAg levels fail to fall within the first 12–24 weeks or to reach <20,000 IU/mL by week 24, PEG IFN therapy is unlikely to be effective and should be discontinued. (Similar observations of HBsAg levels in oral-agent-treated patients are of interest, but of limited clinical relevance, given the very high likelihood of virologic responses during such therapy.)

ENTECAVIR

Entecavir, an oral cyclopentyl guanosine analogue polymerase inhibitor (approved 2005), appears to be the most potent of the HBV antivirals and is just as well tolerated as lamivudine. In a 709-subject clinical trial among HBeAg-reactive patients, oral entecavir, 0.5 mg daily, was compared to lamivudine, 100 mg daily. At 48 weeks, entecavir was superior to lamivudine in suppression of HBV DNA (mean 6.9 vs 5.5 log₁₀ copies/mL), percentage with undetectable HBV DNA (<300 copies/mL by PCR; 67% vs 36%), histologic improvement (≥2-point improvement in necroinflammatory HAI score; 72% vs 62%), and normal ALT (68% vs 60%). The two treatments were indistinguishable in percentage with HBeAg loss (22% vs 20%) and seroconversion (21% vs 18%). Among patients treated with entecavir for 96 weeks, HBV DNA was undetectable cumulatively in 80% (vs 39% for lamivudine), and HBeAg seroconversions had occurred in 31% (vs 26% for lamivudine). After 3–6 years of entecavir, HBeAg seroconversions have been observed in 39–44% and HBsAg loss in 5–6%. Similarly, in a 638-subject clinical trial among HBeAg-negative patients, at week 48, oral entecavir, 0.5 mg daily, was superior to lamivudine, 100 mg daily, in suppression of HBV DNA (mean 5.0 vs 4.5 log₁₀ copies/mL) and in percentage with undetectable HBV DNA (90% vs 72%), histologic improvement (70% vs 61%), and normal ALT (78% vs 71%). No resistance mutations were encountered in previously treatment-naïve, entecavir-treated patients during 96 weeks of therapy, and in a cohort of subjects treated for up to 6 years, resistance emerged in only 1.2%. Entecavir-induced HBeAg seroconversions are as durable as those achieved with other antivirals. Its high barrier to resistance coupled with its high potency renders entecavir a first-line drug for patients with chronic hepatitis B.

Entecavir is also effective against lamivudine-resistant HBV infection. In a trial of 286 lamivudine-resistant patients, entecavir, at a higher daily dose of 1 mg, was superior to lamivudine, as measured at week 48, in achieving suppression of HBV DNA (mean 5.1 vs 0.48 log₁₀ copies/mL), undetectable HBV DNA (72% vs 19%), normal ALT (61% versus 15%), HBeAg loss (10% vs 3%), and HBeAg seroconversion (8% vs 3%). In this population of lamivudine-experienced patients, however, entecavir resistance emerged in 7% at 48 weeks. Although entecavir resistance requires both a YMDD mutation and a second mutation at one of several other sites (e.g., T184A, S202G/I, or M250V), resistance to entecavir in lamivudine-resistant chronic hepatitis B has been recorded to increase progressively to 43% at 4 years and 57% at 6 years; therefore, entecavir is not as attractive a choice (and is not recommended, despite its approval for this indication) as adefovir or tenofovir for patients with lamivudine-resistant hepatitis B.

In clinical trials, entecavir had an excellent safety profile. In addition, on-treatment and posttreatment ALT flares are relatively uncommon and relatively mild in entecavir-treated patients. Doses should be reduced for patients with reduced creatinine clearance. Entecavir does have low-level antiviral activity against HIV and cannot be used as monotherapy to treat HBV infection in HIV/HBV co-infected persons.

TELBIVUDINE

Telbivudine, a cytosine analogue (approved 2006), is similar in efficacy to entecavir but slightly less potent in suppressing HBV DNA (a slightly less profound median 6.4 log₁₀ reduction in HBeAg-reactive disease and a similar 5.2 log₁₀ reduction in HBeAg-negative disease). In its registration trial, telbivudine at an oral daily dose of 600 mg suppressed HBV DNA to <300 copies/mL in 60% of HBeAg-positive and 88% of HBeAg-negative patients, reduced ALT to normal in 77% of HBeAg-positive and 74% of HBeAg-negative patients, and improved histology in 65% of HBeAg-positive and 67% of HBeAg-negative patients. Although resistance to telbivudine (M204I, not M204V, mutations) was less frequent than resistance to lamivudine at the end of 1 year, resistance mutations after 2 years of treatment occurred in up to 22%. Generally well tolerated, telbivudine has been associated with a low frequency of asymptomatic creatine kinase elevations and with a very low frequency of peripheral neuropathy; frequency of administration should be reduced for patients with impaired creatinine clearance. Its excellent potency notwithstanding, the inferior resistance and safety profile of telbivudine has limited its appeal; telbivudine is neither recommended as first-line therapy nor widely used.

TENOFOVIR

TDF, an acyclic nucleotide analogue and potent antiretroviral agent used to treat HIV infection (approved for hepatitis B in 2008), is similar to adefovir but more potent in suppressing HBV DNA and inducing HBeAg responses; it is highly active against both wild-type and lamivudine-resistant HBV and active in patients whose response to adefovir is slow and/or limited. At an oral once-daily dose of 300 mg for 48 weeks, tenofovir suppressed HBV DNA by 6.2 log₁₀ (to undetectable levels [<400 copies/mL] in 76%) in HBeAg-positive patients and by 4.6 log₁₀ (to undetectable levels in 93%) in HBeAg-negative patients; reduced ALT to normal in 68% of HBeAg-positive and 76% of HBeAg-negative patients; and improved histology in 74% of HBeAg-positive and 72% of HBeAg-negative patients. In HBeAg-positive patients, HBeAg seroconversions occurred in 21% by the end of year 1, 27% by year 2, 34% by year 3, and 40% by year 5 of tenofovir treatment; HBsAg loss occurred in 3% by the end of year 1 and 6% at year 2, and 8% by year 5. After 5 years of tenofovir therapy, 87% of patients

experienced histologic improvement, including reduction in fibrosis score (51%) and regression of cirrhosis (71%). The 5-year safety (negligible renal toxicity, in 1%, and mild reduction in bone density, in ~0.5%) and resistance profiles (none recorded through 8 years) of tenofovir are very favorable as well; therefore, tenofovir has supplanted adefovir both as first-line therapy for chronic hepatitis B and as add-on therapy for lamivudine-resistant chronic hepatitis B. Studies of tenofovir and entecavir reviewed in 2015 showed no difference in long-term risks of renal and bone toxicity; however, among patients treated with tenofovir, instances of acute renal failure and of low blood phosphate levels have been reported. Thus, in patients receiving tenofovir, monitoring bone density is not recommended, but periodic (at least annual) monitoring for renal injury is (serum creatinine and phosphate, urine glucose and protein). Frequency of tenofovir administration should be reduced for patients with impaired creatinine clearance.

A comparison of the six antiviral therapies in current use appears in Table 334-3; their relative potencies in suppressing HBV DNA are shown in Fig. 334-1.

COMBINATION THERAPY

Although the combination of lamivudine and PEG IFN suppresses HBV DNA more profoundly during therapy than does monotherapy with either drug alone (and is much less likely to be associated with lamivudine resistance), this combination used for a year is no better than a year of PEG IFN in achieving sustained responses. To date, combinations of oral nucleoside/nucleotide agents have not achieved an enhancement in virologic, serologic, or biochemical efficacy over that achieved by the more potent of the combined drugs given individually. In a 2-year trial of combination entecavir and tenofovir versus entecavir monotherapy, for a small subgroup of patients with very high HBV DNA levels ($\geq 10^8$ IU/mL), a reduction in HBV DNA to <50 IU/mL was higher in the combination group (79% vs 62%); however, no differences in HBeAg responses or any other endpoint were observed between the combination-therapy and monotherapy groups, even in the high-HBV DNA subgroup. On the other hand, combining agents that are not cross-resistant (e.g., lamivudine or entecavir with adefovir or tenofovir) has the potential to reduce the risk or perhaps even to preempt entirely the emergence of drug resistance. In the future, the treatment paradigm may shift from the current approach of sequential monotherapy to preemptive combination therapy, perhaps not for all patients but for subsets (e.g., patients with very high levels of HBV DNA, immunosuppressed patients); however, designing and executing clinical trials that demonstrate superior efficacy and resistance profile of combination therapy over monotherapy with entecavir or tenofovir will remain challenging. Whereas, initially, in clinical studies of adefovir as rescue therapy for lamivudine resistance, *adding* adefovir to lamivudine (combination therapy) was considered a better strategy than replacing lamivudine with adefovir monotherapy, according to the 2016 treatment recommendations of the American Association for the Study of Liver Diseases (AASLD), data to support adding or switching agents are insufficient. Therefore, while sound virologic principles would favor adding as opposed to switching, according to current recommendations involving the more potent first-line agents, entecavir for tenofovir resistance and tenofovir for entecavir resistance, either strategy is acceptable. For patients who already have acquired multidrug resistance (to both nucleoside analogues [lamivudine, entecavir, telbivudine] and nucleotide analogues [adefovir, tenofovir]), treatment with a combination of entecavir and tenofovir has been shown to be highly effective in suppressive HBV DNA and overcoming drug resistance.

NOVEL ANTIVIRALS AND STRATEGIES

In addition to the seven approved antiviral drugs for hepatitis B, emtricitabine, a fluorinated cytosine analogue very similar to lamivudine in structure, efficacy, and resistance profile, offers no advantage over lamivudine. A combination of emtricitabine and tenofovir is approved for the treatment of HIV infection and is an appealing combination therapy for hepatitis B, especially for lamivudine-resistant disease; however, neither emtricitabine nor the combination is approved for hepatitis B. Several initially promising antiviral agents have been abandoned because of toxicity (e.g., clevudine, which was linked to myopathy during its clinical development). As noted above, the current formulation of tenofovir, TDF, has been associated with renal toxicity and loss of bone density, especially in patients with HIV infection, less so in patients with HBV infection. A new formulation, tenofovir alafenamide (TAF), is a prodrug of tenofovir that is metabolized to the active agent in its target organ (the liver for HBV infection); such targeting permits higher dose delivery to the liver with markedly reduced systemic exposure. Studies in patients with chronic hepatitis B treated with 25 mg of TAF or 300 mg of TDF demonstrate comparable virologic efficacy as well as less reduction in bone mineral density and estimated glomerular filtration rate for TAF. Based on its better renal and bone safety profile than TDF, TAF has been approved for HBV infection and provides an alternative to TDF in patients with TDF-associated elevations in serum creatinine and/or reductions in serum phosphorus. Direct-acting antivirals (DAAs) have been very successful in the management of chronic hepatitis B; however, most patients require long-duration, usually indefinite, therapy. Ideally, an approach to achieving “cure” (eradication of HBV infection) with finite-duration therapy would be welcome. Currently, innovative approaches being investigated include viral entry inhibitors, nucleocapsid assembly inhibitors, HBV secretion (HBsAg release) inhibitors, immunomodulators (e.g., toll receptor agonists, T-cell vaccines, programmed cell death [PD-1] blockade, reconstitution of innate immune responses [e.g., retinoic acid-inducible gene-1, RIG-1, HBV mRNA recognition, and activation of innate immune signaling) and of adaptive immune responses, covalently closed circular (ccc) DNA silencing/inhibition/cleavage, RNA interference, and HBx inhibitors. While data supporting several of these unconventional approaches have begun to appear, none has been shown to “cure” hepatitis B, and none is likely to be competitive, unless it can be shown to go beyond current antivirals in achieving recovery (HBsAg seroconversion) from HBV infection. Finally, initial emphasis in the development of antiviral therapy for hepatitis B was placed on monotherapy; whether combination regimens will yield additive or synergistic efficacy remains to be determined.

TREATMENT RECOMMENDATIONS

Several learned societies and groups of expert physicians have issued treatment recommendations for patients with chronic hepatitis B; the most authoritative and updated (and free of financial support by pharmaceutical companies) are those of the AASLD and of the European Association for the Study of the Liver (EASL). Although the recommendations differ slightly, a consensus has emerged on most of the important points (**Table 334-4**). No treatment is recommended or available for inactive “nonreplicative” hepatitis B carriers (undetectable HBeAg with normal ALT and HBV DNA $\leq 10^3$ IU/mL documented serially over time). In patients with detectable HBeAg and HBV DNA levels $> 2 \times 10^4$ IU/mL, treatment is recommended by the AASLD for those with ALT levels above $2 \times$ the upper limit of normal. (The EASL recommends treatment in HBeAg-positive patients for HBV DNA levels $> 2 \times 10^3$ IU/mL and ALT above the upper limit of normal.) For HBeAg-positive patients with ALT $\leq 2 \times$ the upper limit of normal, in whom sustained responses are not likely and who would require multiyear therapy,

antiviral therapy is not recommended currently. This pattern is common during the early decades of life among Asian patients infected at birth; even in this group, therapy would be considered for those >40 years of age, ALT persistently at the high end of the twofold range, and/or with a family history of HCC, especially if the liver biopsy shows moderate to severe necroinflammatory activity or fibrosis. In this group, when, eventually, ALT becomes elevated later in life, antiviral therapy should be instituted. For patients with HBeAg-negative chronic hepatitis B, ALT $>2 \times$ the upper limit of normal (above the upper limit of normal according to EASL), and HBV DNA $>2 \times 10^3$ IU/mL, antiviral therapy is recommended. If HBV DNA is $>2 \times 10^3$ IU/mL and ALT is 1 to $>2 \times$ the upper limit of normal, liver biopsy should be considered to help in arriving at a decision to treat if substantial liver injury is present (treatment in this subset would be recommended according to EASL guidelines, because ALT is elevated). Per current AASLD recommendations, antiviral treatment with oral agents can be stopped after HBeAg seroconversion in noncirrhotics, and the suggested period of consolidation therapy is 12 months with close monitoring for recurrent viremia (monthly $\times 6$, then every 3 months for the rest of a year) after cessation of therapy. For patients with HBeAg-negative chronic hepatitis, the current recommendation with oral agents is for indefinite therapy; although sufficient data are lacking, stopping therapy in this group can be considered after HBsAg loss.

For patients with compensated cirrhosis, because antiviral therapy has been shown to retard clinical progression, treatment is recommended regardless of HBeAg status and ALT as long as HBV DNA is detectable at $>2 \times 10^3$ IU/mL (detectable at any level according to the EASL); monitoring without therapy is recommended for those with HBV DNA $<2 \times 10^3$ IU/mL, unless ALT is elevated. For patients with decompensated cirrhosis, treatment is recommended regardless of serologic and biochemical status, as long as HBV DNA is detectable. Patients with decompensated cirrhosis should be evaluated as candidates for liver transplantation.

Among the seven available drugs for hepatitis B, PEG IFN has supplanted standard IFN, entecavir has supplanted lamivudine, and tenofovir has supplanted adefovir. PEG IFN, entecavir, or tenofovir is recommended as first-line therapy (Table 334-3). PEG IFN requires finite-duration therapy, achieves the highest rate of HBeAg responses after a year of therapy, and does not support viral mutations, but it requires subcutaneous injections and is associated with inconvenience, more intensive clinical and laboratory monitoring, and intolerability. Oral nucleoside analogues require long-term therapy in most patients, and when used alone, lamivudine and telbivudine foster the emergence of viral mutations, adefovir somewhat less so, and entecavir (except in lamivudine-experienced patients) and tenofovir rarely at all. Oral agents do not require injections or cumbersome laboratory monitoring, are very well tolerated, lead to improved histology in 50–90% of patients, suppress HBV DNA more profoundly than PEG IFN, and are effective even in patients who fail to respond to IFN-based therapy. Although oral agents are less likely to result in HBeAg responses during the first year of therapy, as compared to PEG IFN, treatment with oral agents tends to be extended beyond the first year and, by the end of the second year, yields HBeAg responses (and even HBsAg responses) comparable in frequency to those achieved after 1 year of PEG IFN (and without the associated side effects) (Table 334-5). In a 2016 systematic review of 1716 patients involved in 25 clinical trials, responses after oral-agent therapy were found to be durable. Among patients with HBeAg-reactive chronic hepatitis B, the pooled rates of durable HBeAg seroconversions maintained after cessation of nucleoside/nucleotide analogue therapy (including all the oral agents) were 92% and 88% at posttreatment months 12 and 24, respectively, unaffected by the

duration of post-HBeAg-response consolidation therapy (>6 months in all studies evaluated); the pooled rate of durable biochemical remission after therapy in this population was 76%. Even for HBeAg-negative chronic hepatitis B, for which most authorities recommend indefinite therapy, pooled rates of virologic remissions maintained after cessation of oral-agent therapy were 44%, 31%, and 30% at posttreatment months 12, 24, and 36, and the pooled rate of durable biochemical remission in this population was 57%.

Although adefovir and tenofovir are safe, renal monitoring (e.g., serum creatinine and phosphate, urine glucose and protein) is recommended. Substantial experience with lamivudine during pregnancy (see above) has identified no teratogenicity; although widely used during pregnancy, lamivudine remains classified as pregnancy category C. Although IFNs do not appear to cause congenital anomalies, these have antiproliferative properties and should be avoided during pregnancy. Adefovir during pregnancy has not been associated with birth defects; however, the risk of spontaneous abortion may be increased, and adefovir is categorized as pregnancy category C. Data on the safety of entecavir during pregnancy have not been published (pregnancy category C). Sufficient data in animals and limited data in humans suggest that telbivudine and tenofovir (both pregnancy category B) can be used safely during pregnancy; however, telbivudine is not an acceptable first-line drug. In general, then, except for lamivudine and tenofovir, and until additional data become available, the other antivirals for hepatitis B should be avoided or used with extreme caution during pregnancy.

For children aged 2 to <18 with HBeAg-reactive hepatitis B (most children will be HBeAg-reactive; no studies have been done in children with HBeAg-negative chronic hepatitis B), treatment is recommended if HBV DNA is detectable and ALT levels are elevated, but not if ALT levels are normal. Each of the available drugs, except telbivudine, is approved for different childhood age groups (standard IFN α -2b age ≥ 1 year; PEG IFN α -2a age ≥ 5 years [approved for hepatitis C, not B, but can be used in hepatitis B]; lamivudine and entecavir age ≥ 2 years; adefovir and tenofovir age ≥ 12 years). Package inserts should be consulted for childhood doses.

As noted above, some physicians prefer to begin with PEG IFN, while other physicians and patients prefer oral agents as first-line therapy. For patients with decompensated cirrhosis, the emergence of resistance can result in further deterioration and loss of antiviral effectiveness. Therefore, in this patient subset, the threshold for relying on therapy with a very favorable resistance profile (e.g., entecavir or tenofovir) or on combination therapy is low. PEG IFN should not be used in patients with compensated or decompensated cirrhosis.

For patients with end-stage chronic hepatitis B who undergo liver transplantation, reinfection of the new liver is almost universal in the absence of antiviral therapy. The majority of patients become high-level viremic carriers with minimal liver injury. Before the availability of antiviral therapy, an unpredictable proportion experienced severe hepatitis B-related liver injury, sometimes a fulminant-like hepatitis and sometimes a rapid recapitulation of the original severe chronic hepatitis B (**Chap. 332**). Currently, however, prevention of recurrent hepatitis B after liver transplantation has been achieved definitively by *combining* hepatitis B immune globulin with one of the low-resistance oral nucleoside (entecavir) or nucleotide analogues (tenofovir) (**Chap. 338**); preliminary data suggest that the newer, more potent, and less resistance-prone oral agents may be used instead of hepatitis B immune globulin for posttransplantation therapy. In patients documented at the time of liver transplantation to have undetectable HBV DNA in serum and cccDNA in the liver (i.e., with low risk for recurrence of HBV infection), a preliminary clinical trial suggested that, after patients received 5 years of

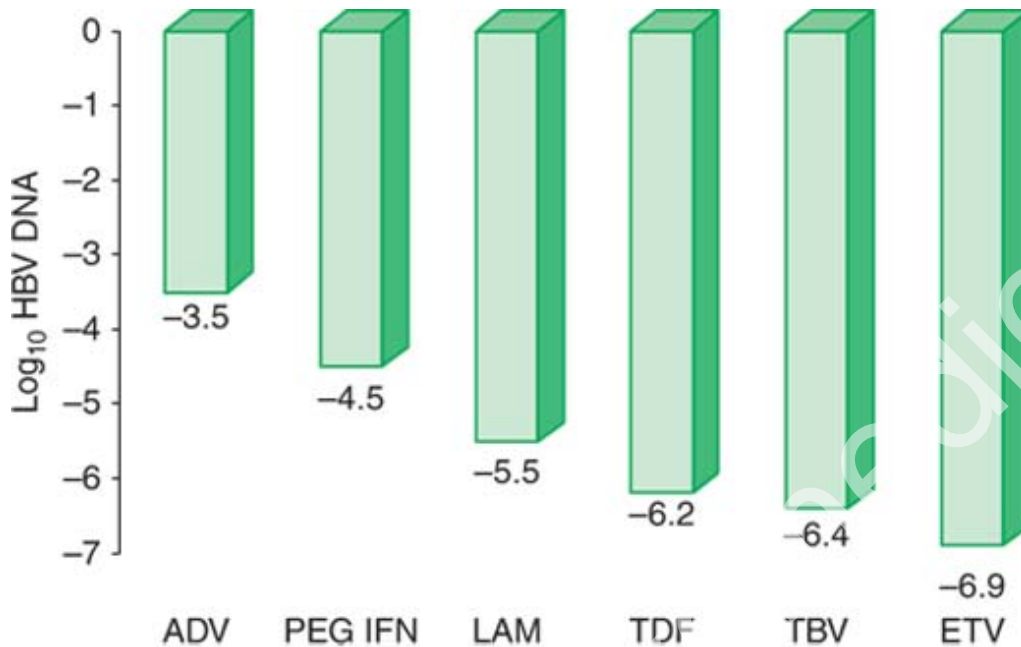
combined therapy, both hepatitis B immune globulin and oral-agent therapy can be withdrawn sequentially (over two 6-month periods) with a success rate, as monitored over a median of 6 years postwithdrawal, of 90% and an anti-HBs seroconversion rate of 60% (some with transient reappearance of HBV DNA and/or HBsAg).

Patients with HBV-HIV co-infection can have progressive HBV-associated liver disease and, occasionally, a severe exacerbation of hepatitis B resulting from immunologic reconstitution following ART. Lamivudine should never be used as monotherapy in patients with HBV-HIV infection because HIV resistance emerges rapidly to both viruses. Adefovir has been used successfully to treat chronic hepatitis B in HBV-HIV co-infected patients but is no longer considered a first-line agent for HBV. Entecavir has low-level activity against HIV and can result in selection of HIV resistance; therefore, it should be avoided in HBV-HIV co-infection. Tenofovir and the combination of tenofovir and emtricitabine in one pill are approved therapies for HIV and represent excellent choices for treating HBV infection in HBV-HIV co-infected patients. Generally, even for HBV-HIV co-infected patients who do not yet meet treatment criteria for HIV infection, treating for both HBV and HIV is recommended.

Patients with chronic hepatitis B who undergo cytotoxic chemotherapy for treatment of malignancies as well as patients treated with immunosuppressive, anticytokine, or antitumor necrosis factor therapies (the risk varies, from highest [e.g., B-cell-depleting agents, anthracycline derivatives, moderate/high-dose corticosteroids for ≥ 4 weeks] to moderate [e.g., tumor necrosis factor alpha inhibitors, cytokine or integrin inhibitors, tyrosine kinase inhibitors, low-dose corticosteroids for ≥ 4 weeks], to lowest [e.g., immunosuppressive agents like methotrexate and azathioprine, intraarticular corticosteroids, any dose of corticosteroids for ≤ 1 week]) experience enhanced HBV replication and viral expression on hepatocyte membranes during chemotherapy coupled with suppression of cellular immunity. When chemotherapy is withdrawn, such patients are at risk for reactivation of hepatitis B, often severe and occasionally fatal. Such rebound reactivation represents restoration of cytolytic T cell function against a target organ enriched in HBV expression. Preemptive treatment with the first of the oral HBV antivirals, lamivudine, prior to the initiation of chemotherapy was shown to reduce the risk of such reactivation substantially; treating *after* reactivation has occurred is less effective. The newer, more potent oral antiviral agents, entecavir and tenofovir, which are even more effective in preventing hepatitis B reactivation and with a lower risk of antiviral drug resistance, are preferred. The optimal duration of antiviral therapy after completion of chemotherapy is not known, but a suggested approach is 6 months (12 months for B-cell-depleting agents) for inactive hepatitis B carriers and longer-duration therapy in patients with baseline HBV DNA levels $> 2 \times 10^3$ IU/mL, until standard clinical endpoints are met (Table 334-4). Such chemotherapy-associated reactivation of hepatitis B is common (4–68%, median 25%, in a meta-analysis) in persons with ongoing HBV infection (HBsAg-reactive); however, such reactivation can occur albeit less commonly in persons who have cleared HBsAg, but express anti-HBc (moderate risk, $< 10\%$) and rarely ($< 5\%$) even in persons with serologic evidence of recovery from HBV infection (anti-HBs-reactive, anti-HBc-reactive). Therefore, most authorities (e.g., Centers for Disease Control and Prevention; AASLD; American Gastroenterological Association; EASL) recommend HBsAg and anti-HBc (\pm anti-HBs) screening of all patients undergoing such chemotherapy and preemptive antiviral prophylaxis for HBsAg-reactive persons and close on-therapy monitoring of anti-HBc-reactive/anti-HBs-reactive persons with treatment if and when reactivation occurs.

FIGURE 334-1

Relative potency of antiviral drugs for hepatitis B, as reflected by median \log_{10} HBV DNA reduction in HBeAg-positive chronic hepatitis B. These data are from individual reports of large, randomized controlled registration trials that were the basis for approval of the drugs. In most instances, these data do not represent direct comparisons among the drugs, because study populations were different, baseline patient variables were not always uniform, and the sensitivity and dynamic range of the HBV DNA assays used in the trials varied. ADV, adefovir dipivoxil; ETV, entecavir; LAM, lamivudine; PEG IFN, pegylated interferon $\alpha 2a$; TBV, telbivudine; TDF, tenofovir disoproxil fumarate.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

TABLE 334-3

Comparison of Pegylated Interferon (PEG IFN), Lamivudine, Adefovir, Entecavir, Telbivudine, and Tenofovir Therapy for Chronic Hepatitis B^a

FEATURE	PEG IFN ^b	LAMIVUDINE	ADEFOVIR	ENTECAVIR	TELBIVUDINE	TENOFOVIR
Route of administration	Subcutaneous injection	Oral	Oral	Oral	Oral	Oral
Duration of therapy ^c	48–52 weeks	≥52 weeks	≥48 weeks	≥48 weeks	≥52 weeks	≥48 weeks
Tolerability	Poorly tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended
HBeAg seroconversion 1 yr Rx >1 yr Rx	18–20% NA	16–21% up to 50% @ 5 yrs	12% 43% @ 3 yrs ^d	21% 31% @ 2 yrs 44% @ 6 yrs	22% 30% @ 2 yrs	21% 40% @ 5 yrs
Log ₁₀ HBV DNA reduction (mean copies/mL) HBeAg-reactive HBeAg-negative	4.5 4.1	5.5 4.4–4.7	median 3.5–5 median 3.5– 3.9	6.9 5.0	6.4 5.2	6.2 4.6

FEATURE	PEG IFN ^b	LAMIVUDINE	ADEFOVIR	ENTECAVIR	TELBIVUDINE	TENOFOVIR
HBV DNA PCR negative (<300–400 copies/mL; <1000 copies/mL for adefovir) at end of yr 1 HBeAg-reactive HBeAg-negative	10–25% 63%	36–44% 60–73%	13–21% 48–77%	67% (91% @ 4 yrs) 90%	60% 88%	76% 93%
ALT normalization at end of yr 1 HBeAg-reactive HBeAg-negative	39% 34–38%	41–75% 62–79%	48–61% 48–77%	68% 78%	77% 74%	68% 76%
HBsAg loss yr 1 >yr 1	3–4% 12% 5 yr after 1 yr of Rx	≤1% No data	0% 5% at yr 5	2% 6% at yr 6	<1% No data	3% 8% at yr 5
Histologic improvement (≥2 point reduction in HAI) at yr 1 HBeAg-reactive HBeAg-negative	38% 6 months after 48% 6 months after	49–62% 61–66%	53–68% 64%	72% 70%	65% 67%	74% 72%
Viral resistance	None	15–30% @ 1 yr 70% @ 5 yrs	None @ 1 yr 29% @ 5 yrs	≤1% @ 1 yr ^e 1.2% @ 6 yrs ^e	Up to 5% @ yr 1 Up to 22% @ yr 2	0% @ yr 1 0% through yr 8

FEATURE	PEG IFN ^b	LAMIVUDINE	ADEFOVIR	ENTECAVIR	TELBIVUDINE	TENOFOVIR
Pregnancy category	C	C ^f	C	C	B	B
Cost (US\$) for 1 yr	~\$18,000	~\$2,500	~\$6,500	~\$8,700 ^g	~\$6,000	~\$6,000

^aGenerally, these comparisons are based on data on each drug tested individually versus placebo in registration clinical trials; because, with rare exception, these comparisons are not based on head-to-head testing of these drugs, relative advantages and disadvantages should be interpreted cautiously. ^bAlthough standard interferon α administered daily or three times a week is approved as therapy for chronic hepatitis B, it has been supplanted by PEG IFN, which is administered once a week and is more effective. Standard interferon has no advantages over PEG IFN. ^cDuration of therapy in clinical efficacy trials; use in clinical practice may vary. ^dBecause of a computer-generated randomization error that resulted in misallocation of drug versus placebo during the second year of clinical trial treatment, the frequency of HBeAg seroconversion beyond the first year is an estimate (Kaplan-Meier analysis) based on the small subset in whom adefovir was administered correctly. ^e7% during a year of therapy (43% at year 4) in lamivudine-resistant patients.

^fDespite its Class C designation, lamivudine has an extensive pregnancy safety record in women with HIV/AIDS.

^gApproximately \$17,400 for lamivudine-refractory patients.

Abbreviations: ALT, alanine aminotransferase; HAI, histologic activity index; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; NA, not applicable; PEG IFN, pegylated interferon; PCR, polymerase chain reaction; Rx, therapy; yr, year.

TABLE 334-4

Recommendations for Treatment of Chronic Hepatitis B^a

HBeAg STATUS	CLINICAL	HBV DNA (IU/mL)	ALT	RECOMMENDATION
HBeAg-reactive	^b Chronic hepatitis Cirrhosis compensated Cirrhosis decompensated	$>2 \times 10^4$ $>2 \times 10^{4d}$ $>2 \times 10^3$ $<2 \times 10^3$ Detectable Undetectable	$\leq 2 \times$ ULN ^{c,d} $>2 \times$ ULN ^d < or > ULN >ULN < or > ULN < or > ULN	No treatment; monitor. In patients >40, with family history of hepatocellular carcinoma, and/or ALT persistently at the high end of the twofold range, liver biopsy may help in decision to treat Treat ^e Treat ^e with oral agents, not PEG IFN Consider treatment ^f Treat ^e with oral agents ^g , not PEG IFN; refer for liver transplantation Observe; refer for liver transplantation
HBeAg-negative	^b Chronic hepatitis Chronic hepatitis Cirrhosis compensated Cirrhosis decompensated	$\leq 2 \times 10^3$ $>10^3$ $>10^4$ $>2 \times 10^3$ $<2 \times 10^3$ Detectable Undetectable	\leq ULN 1 to >2 \times ULN ^d $>2 \times$ ULN ^d < or > ULN >ULN < or > ULN < or > ULN	Inactive carrier; treatment not necessary Consider liver biopsy; treat ^h if biopsy shows moderate to severe inflammation or fibrosis Treat ^{h,i} Treat ^e with oral agents, not PEG IFN Consider treatment ^f Treat ^h with oral agents ^g , not PEG IFN; refer for liver transplantation Observe; refer for liver transplantation

^aBased on practice guidelines of the American Association for the Study of Liver Diseases (AASLD). Except as indicated in footnotes, these guidelines are similar to those issued by the European Association for the Study of the Liver (EASL).

^bLiver disease tends to be mild or inactive clinically; most such patients do not undergo liver biopsy. ^cThis pattern is common during early decades of life in Asian patients infected at birth. ^dAccording to the EASL guidelines, treat if HBV DNA is $>2 \times 10^3$ IU/mL and ALT $>$ ULN. ^eOne of the potent oral drugs with a high barrier to resistance (entecavir or tenofovir) or PEG IFN can be used as first-line therapy (see text). These oral agents, but not PEG IFN, should be used for

interferon-refractory/intolerant and immunocompromised patients. PEG IFN is administered weekly by subcutaneous injection for a year; the oral agents are administered daily for at least a year and continued indefinitely or until at least 6 months after HBeAg seroconversion. ^fAccording to EASL guidelines, patients with compensated cirrhosis and detectable HBV DNA at any level, even with normal ALT, are candidates for therapy. Most authorities would treat indefinitely, even in HBeAg-positive disease after HBeAg seroconversion. ^gBecause the emergence of resistance can lead to loss of antiviral benefit and further deterioration in decompensated cirrhosis, a low-resistance regimen is recommended—entecavir or tenofovir monotherapy or combination therapy with the more resistance-prone lamivudine (or telbivudine) plus adefovir. Therapy should be instituted urgently. ^hBecause HBeAg seroconversion is not an option, the goal of therapy is to suppress HBV DNA and maintain a normal ALT. PEG IFN is administered by subcutaneous injection weekly for a year; caution is warranted in relying on a 6-month posttreatment interval to define a sustained response, because the majority of such responses are lost thereafter. Oral agents, entecavir or tenofovir, are administered daily, usually indefinitely or until, as very rarely occurs, virologic and biochemical responses are accompanied by HBsAg seroconversion. ⁱFor older patients and those with advanced fibrosis, consider lowering the HBV DNA threshold to $>2 \times 10^3$ IU/mL.

Abbreviations: AASLD, American Association for the Study of Liver Diseases; ALT, alanine aminotransferase; EASL, European Association for the Study of the Liver; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; PEG IFN, pegylated interferon; ULN, upper limit of normal.

TABLE 334-5

Pegylated Interferon Versus Oral Nucleoside Analogues for the Treatment of Chronic Hepatitis B

	PEG IFN	NUCLEOSIDE ANALOGUES
Administration	Weekly injection	Daily, orally
Tolerability	Poorly tolerated, intensive monitoring	Well tolerated, limited monitoring
Duration of therapy	Finite 48 weeks	≥1 year, indefinite in most patients
Maximum mean HBV DNA suppression	4.5 log ₁₀	6.9 log ₁₀
Effective in high-level HBV DNA (≥10 ⁹ IU/mL)	No	Yes
HBeAg seroconversion During 1 year of therapy During >1 year of therapy	~30% Not applicable	~20% 30% (year 2) to up to 50% (year 5)
HBeAg-negative posttreatment HBV DNA suppression	17% @ 5 years	7% @ 4 years (lamivudine)
HBsAg loss During 1 year of therapy During >1 year of therapy After 1 year of therapy–HBeAg-negative	3–4% Not applicable 12% @ 5 years	0–3% 3–8% @ 5 years of therapy 3.5% @ 5 years
Antiviral resistance	None	Lamivudine: ~30% year 1, ~70% year 5 Adefovir: 0% year 1, ~30% year 5 Telbivudine: up to 4% year 1, 22% year 2 Entecavir: ≤1.2% through year 6 Tenofovir: 0% through year 8
Use in cirrhosis, transplantation, immunosuppressed	No	Yes

	PEG IFN	NUCLEOSIDE ANALOGUES
Cost, 1 year of therapy	++++	+ to ++

Abbreviations: HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; PEG IFN, pegylated interferon.

CHRONIC HEPATITIS D (DELTA HEPATITIS)

Chronic hepatitis D virus (HDV) may follow acute co-infection with HBV but at a rate no higher than the rate of chronicity of acute hepatitis B. That is, although HDV co-infection can increase the severity of acute hepatitis B, HDV does not increase the likelihood of progression to chronic hepatitis B. When, however, HDV superinfection occurs in a person who is already chronically infected with HBV, long-term HDV infection is the rule, and a worsening of the liver disease is the expected consequence. Except for severity, chronic hepatitis B plus D has similar clinical and laboratory features to those seen in chronic hepatitis B alone. Relatively severe and progressive chronic hepatitis, with or without cirrhosis, is the rule, and mild chronic hepatitis is the exception. Occasionally, however, mild hepatitis or even, rarely, inactive carriage occurs in patients with chronic hepatitis B plus D, and the disease may become indolent after several years of infection. A distinguishing serologic feature of chronic hepatitis D is the presence in the circulation of antibodies to liver-kidney microsomes (anti-LKM); however, the anti-LKM seen in hepatitis D, anti-LKM3, are directed against uridine diphosphate glucuronosyltransferase and are distinct from anti-LKM1 seen in patients with autoimmune hepatitis and in a subset of patients with chronic hepatitis C (see below). **The clinical and laboratory features of chronic HDV infection are summarized in Chap. 332.**

TREATMENT

TREATMENT: Chronic Hepatitis D

Management is not well defined, and the host cellular RNA polymerase upon which HDV replication depends cannot be targeted by conventional antiviral agents. Glucocorticoids are ineffective and are not used. Preliminary experimental trials of IFN- α suggested that conventional doses and durations of therapy lower levels of HDV RNA and aminotransferase activity only transiently during treatment but have no impact on the natural history of the disease. In contrast, high-dose IFN- α (9 million units three times a week) for 12 months was reported to be associated with a sustained loss of HDV replication and clinical improvement in up to 50% of patients. Moreover, in anecdotal reports, the beneficial impact of treatment has been observed to persist for 15 years and to be associated with a reduction in grade of hepatic necrosis and inflammation, reversion of advanced fibrosis (improved stage), and clearance of HDV RNA in some patients. A suggested approach to therapy has been high-dose, long-term IFN for at least a year and, in responders, extension of therapy until HDV RNA and HBsAg clearance; however, extension of therapy to a second year provided no advantage, and sustained responses after completion of therapy have been rare. PEG IFN has also been shown to be more

effective in the treatment of chronic hepatitis D (e.g., after 48 weeks of therapy, associated with undetectable HDV RNA, durable for at least 24 posttreatment weeks, in a quarter to a half of patients) and is a more convenient replacement for standard IFN; however, loss of virologic responses (reappearance of HDV RNA) was observed during long-term (median 4.5-year) monitoring in over half of initial, 24-week-posttreatment responders. Even extending PEG IFN therapy for 5 years and driving treatment doses up to 270 µg weekly (of PEG IFN- α2a), as reported in a small trial among 13 patients, while achieving serologic, virologic, histologic, biochemical, and clinical improvement, yielded sustained virologic responses (SVRs) in only 3 patients (58–246 weeks of posttreatment observation). None of the nucleoside analogue antiviral agents for hepatitis B is effective in hepatitis D, and adding oral nucleoside agents to PEG IFN is no more effective than PEG IFN monotherapy. While recommended, PEG IFN therapy is far from satisfactory. Preliminary trials have been performed with an oral prenylation inhibitor, lonafarnib, and with an inhibitor of HBV/HDV viral entry into hepatocytes, myrcludex B. Prenylation, the posttranslational covalent addition of the prenyl lipid farnesyl to large HDV antigen, is required for this HDV protein to interact and form secreted viral particles with HBsAg. In 14 patients treated twice daily for 28 days with 100 or 200 mg of lonafarnib, HDV RNA fell by 0.73 log₁₀ IU/mL and 1.54 log₁₀ IU/mL, respectively, before rebounding after completion of therapy. Hepatitis B virus entry into hepatocytes requires the binding of the myristolated N-terminal pre-S1 peptide of large HBsAg to sodium taurocholate co-transporting peptide, the functional receptor for HBV into hepatocytes. The application of myrcludex B, a synthetic homologous myristolated lipopeptide that competes for binding with HBsAg, was reported in a study of 24 patients (with a baseline mean of 4.1–4.2 log₁₀ copies/mL of HDV RNA) randomized to 24 weeks of treatment with myrcludex B (2 mg daily subcutaneously) as monotherapy or combined with PEG IFN compared to PEG IFN alone. A reduction in HDV RNA occurred in all three groups, by 1.67 log₁₀ copies/mL (in two of eight patients RNA became undetectable), 2.59 log₁₀ copies/mL (in five of eight patients RNA became undetectable), and 2.17 log₁₀ copies/mL (in two of eight patients RNA became undetectable), respectively. No change occurred, however, in the level of HBsAg, which would have been expected. In these two exploratory brief-duration trials, sustained responses were not achieved, and toxicities were encountered (e.g., intermittent vomiting and weight loss [lonafarnib] and transient amylase and lipase elevations [myrcludex B]); however, from these proof-of-principle trials, potentially, more definitive and larger-scale studies will follow.

In patients with end-stage liver disease secondary to chronic hepatitis D, liver transplantation has been effective. If hepatitis D recurs in the new liver without the expression of hepatitis B (an unusual serologic profile in immunocompetent persons but common in transplant patients), liver injury is limited. In fact, the outcome of transplantation for chronic hepatitis D is superior to that for chronic hepatitis B; in such patients, combination hepatitis B [immune globulin](#) and nucleoside analogue therapy for hepatitis B is indicated ([Chap. 338](#)).

CHRONIC HEPATITIS C

Regardless of the epidemiologic mode of acquisition of hepatitis C virus (HCV) infection, chronic hepatitis follows acute hepatitis C in 50–70% of cases; chronic infection is common even in those with a return to normal in aminotransferase levels after acute hepatitis C, adding up to an 85% likelihood of chronic HCV infection after

acute hepatitis C. Few clues had emerged to explain host differences associated with chronic infection until recently, when variation in a single nucleotide polymorphism (SNP) on chromosome 19, *IL28B* (which codes for IFN- λ 3), was identified that distinguished between responders and nonresponders to IFN-based antiviral therapy (see below). The same variants correlated with spontaneous resolution after acute infection: 53% in genotype C/C, 30% in genotype C/T, but only 23% in genotype T/T. The association with HCV clearance after acute infection is even stronger when *IL28B* haplotype is combined with haplotype G/G of a SNP near human leukocyte antigen (HLA) Class II *DBQ1*03:01*.

In patients with chronic hepatitis C followed for 20 years, progression to cirrhosis occurs in about 20–25%. Such is the case even for patients with relatively clinically mild chronic hepatitis, including those without symptoms, with only modest elevations of aminotransferase activity, and with mild chronic hepatitis on liver biopsy. Even in cohorts of well compensated patients with chronic hepatitis C referred for clinical research trials (no complications of chronic liver disease and with normal hepatic synthetic function), the prevalence of cirrhosis may be as high as 50%. Most cases of hepatitis C are identified initially in asymptomatic patients who have no history of acute hepatitis C (e.g., those discovered while attempting to donate blood, while undergoing lab testing as part of an application for life insurance, or as a result of routine laboratory tests). The source of HCV infection in many of these cases is not defined, although a long-forgotten percutaneous exposure (e.g., injection drug use) in the remote past can be elicited in a substantial proportion and probably accounts for most infections; most of these infections were acquired in the 1960s and 1970s, coming to clinical attention decades later.

Approximately one-third of patients with chronic hepatitis C have normal or near-normal aminotransferase activity; although one-third to one-half of these patients have chronic hepatitis on liver biopsy, the grade of liver injury and stage of fibrosis tend to be mild in the vast majority. In some cases, more severe liver injury has been reported—even, rarely, cirrhosis, most likely the result of previous histologic activity. Among patients with persistent normal aminotransferase activity sustained over ≥ 5 –10 years, histologic progression has been shown to be rare; however, approximately one-fourth of patients with normal aminotransferase activity experience subsequent aminotransferase elevations, and histologic injury can be progressive once abnormal biochemical activity resumes. Therefore, continued clinical monitoring and antiviral therapy are indicated, even for patients with normal aminotransferase activity.

Despite this substantial rate of progression of chronic hepatitis C, and despite the fact that liver failure can result from end-stage chronic hepatitis C, the long-term prognosis over 1–2 decades for chronic hepatitis C in a majority of patients is relatively benign. Mortality >10 –20 years among patients with transfusion-associated chronic hepatitis C has been shown not to differ from mortality in a matched population of transfused patients in whom hepatitis C did not develop. Although death in the hepatitis group is more likely to result from liver failure, and although hepatic decompensation may occur in $\sim 15\%$ of such patients over the course of a decade, the majority (almost 60%) of patients remain asymptomatic and well compensated, with no clinical sequelae of chronic liver disease. Overall, chronic hepatitis C tends to be very slowly and insidiously progressive, if at all, in the vast majority of patients, whereas in approximately one-fourth of cases, chronic hepatitis C will progress eventually to end-stage cirrhosis. In fact, because HCV infection is so prevalent, and because a proportion of patients progress inexorably to end-stage liver disease, hepatitis C is the most frequent indication for liver

transplantation (**Chap. 338**). In the United States, hepatitis C accounts for up to 40% of all chronic liver disease; as of 2007, mortality caused by hepatitis C surpassed that associated with HIV/AIDS, and as of 2012, reported deaths caused by hepatitis C surpassed those associated with all other notifiable infectious diseases (HIV, tuberculosis, hepatitis B, and 57 other infectious diseases). Moreover, because the prevalence of HCV infection is so much higher in the “baby boomer” cohort born between 1945 and 1965, three-quarters of the mortality associated with hepatitis C occurs in this age cohort. Referral bias may account for the more severe outcomes described in cohorts of patients reported from tertiary care centers (20-year progression of $\geq 20\%$) versus the more benign outcomes in cohorts of patients monitored from initial blood-product-associated acute hepatitis or identified in community settings (20-year progression of only 4–7%). Still unexplained, however, are the wide ranges in reported progression to cirrhosis, from 2% over 17 years in a population of Irish women with hepatitis C infection acquired from contaminated anti-D immune globulin to 30% over ≤ 11 years in recipients of contaminated intravenous immune globulin.

Progression of liver disease in patients with chronic hepatitis C has been reported to be more likely in patients with older age, longer duration of infection, advanced histologic stage and grade, more complex HCV quasispecies diversity, increased hepatic iron, concomitant other liver disorders (alcoholic liver disease, chronic hepatitis B, hemochromatosis, α_1 antitrypsin deficiency, and steatohepatitis), HIV infection, and obesity. Among these variables, however, duration of infection appears to be one of the most important, and some of the others probably reflect disease duration to some extent (e.g., quasispecies diversity, hepatic iron accumulation). No other epidemiologic or clinical features of chronic hepatitis C (e.g., severity of acute hepatitis, level of aminotransferase activity, level of HCV RNA, presence or absence of jaundice during acute hepatitis) are predictive of eventual outcome. Despite the relatively benign nature of chronic hepatitis C over time in many patients, cirrhosis following chronic hepatitis C has been associated with the late development, after several decades, of HCC (**Chap. 78**); the annual rate of HCC in cirrhotic patients with hepatitis C is 1–4%, occurring primarily in patients who have had HCV infection for 30 years or more.

Perhaps the best prognostic indicator in chronic hepatitis C is liver histology; the rate of hepatic fibrosis may be slow, moderate, or rapid. Patients with mild necrosis and inflammation as well as those with limited fibrosis have an excellent prognosis and limited progression to cirrhosis. In contrast, among patients with moderate to severe necroinflammatory activity or fibrosis, including septal or bridging fibrosis, progression to cirrhosis is highly likely over the course of 10–20 years. The pace of fibrosis progression may be accelerated by such factors as concomitant HIV infection, other causes of liver disease, excessive alcohol use, and hepatic steatosis. Among patients with compensated cirrhosis associated with hepatitis C, the 10-year survival rate is close to 80%; mortality occurs at a rate of 2–6% per year; decompensation at a rate of 4–5% per year; and, as noted above, HCC at a rate of 1–4% per year. Estimates of the natural history of chronic hepatitis C have been made, based on data available on the prevalence of HCV infection in the US population and on the rate of disease progression. Weighted primarily by the concentration of chronic hepatitis C in the baby boomer generation, the peak prevalence was estimated to have occurred in 2015. The calculated frequency of cirrhosis in US patients with hepatitis C was 5% in 1990, 25% in 2010, and is projected to be 37% in 2020. Estimated peak mortality has been predicted to occur in 2032. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis C appears in Chap. 332.**

Clinical features of chronic hepatitis C are similar to those described above for chronic hepatitis B. Generally, fatigue is the most common symptom; jaundice is rare. Immune complex–mediated extrahepatic complications of chronic hepatitis C are less common than in chronic hepatitis B (despite the fact that assays for immune complexes are often positive in patients with chronic hepatitis C), with the exception of essential mixed cryoglobulinemia (**Chap. 332**), which is linked to cutaneous vasculitis and membranoproliferative glomerulonephritis as well as lymphoproliferative disorders such as B-cell lymphoma and unexplained monoclonal gammopathy. In addition, chronic hepatitis C has been associated with extrahepatic complications unrelated to immune-complex injury. These include Sjögren’s syndrome, lichen planus, porphyria cutanea tarda, type 2 diabetes mellitus, and the metabolic syndrome (including insulin resistance and steatohepatitis).

Laboratory features of chronic hepatitis C are similar to those in patients with chronic hepatitis B, but aminotransferase levels tend to fluctuate more (the characteristic episodic pattern of aminotransferase activity) and to be lower, especially in patients with long-standing disease. An interesting and occasionally confusing finding in patients with chronic hepatitis C is the presence of autoantibodies. Rarely, patients with autoimmune hepatitis (see below) and hyperglobulinemia have false-positive immunoassays for anti-HCV. On the other hand, some patients with serologically confirmable chronic hepatitis C have circulating anti-LKM. These antibodies are anti-LKM1, as seen in patients with autoimmune hepatitis type 2 (see below), and are directed against a 33-amino-acid sequence of cytochrome P450 IID6. The occurrence of anti-LKM1 in some patients with chronic hepatitis C may result from the partial sequence homology between the epitope recognized by anti-LKM1 and two segments of the HCV polyprotein. In addition, the presence of this autoantibody in some patients with chronic hepatitis C suggests that autoimmunity may be playing a role in the pathogenesis of chronic hepatitis C.

Histopathologic features of chronic hepatitis C, especially those that distinguish hepatitis C from hepatitis B, are described in [Chap. 332](#).

TREATMENT

TREATMENT: Chronic Hepatitis C

Therapy for chronic hepatitis C has evolved substantially in the 25 years since IFN- α was introduced for this indication in 1991. The therapeutic armamentarium grew to include PEG IFN with [ribavirin](#) and, then, in 2011, the introduction of the first protease inhibitors, telaprevir and boceprevir, used in combination with PEG IFN and [ribavirin](#) in patients with HCV genotype 1. The field of antiviral therapy for hepatitis C was transformed beginning in 2013, with the approval of the first nucleoside analogue, [sofosbuvir](#). As of 2016, no fewer than six, all-oral, highly effective (>95%), low-resistance, well tolerated, short-duration (usually 12 weeks) combination regimens of DAA drugs are available. The remarkable historical evolution of antiviral therapy for hepatitis C is instructive.

THE INTERFEON ERA (1991–2011)

IFN-based therapy has been supplanted by DAA agents introduced in the second decade of the twenty-first century; however, many important lessons about antiviral therapy for chronic hepatitis C were learned from

the experience with IFN-based treatment, and many of the limitations of—and disparities in responsiveness to—IFN-based therapy have been overcome by current-generation DAA treatments. When first approved, IFN- α was administered via subcutaneous injection three times a week for 6 months but achieved an SVR (**Fig. 334-2**) (defined then as a reduction of HCV RNA to undetectable levels by PCR when measured ≥ 24 weeks after completion of therapy) $< 10\%$. Doubling the duration of therapy—but not increasing the dose or changing IFN preparations—increased the SVR rate to $\sim 20\%$, and addition to the regimen of daily **ribavirin**, an oral guanosine nucleoside, increased the SVR rate to 40%. When used alone, **ribavirin** is ineffective and does not reduce HCV RNA levels appreciably, but **ribavirin** enhances the efficacy of IFN by reducing the likelihood of virologic relapse after the achievement of an end-treatment response (**Fig. 334-2**) (response measured during, and maintained to the end of, treatment). Proposed mechanisms to explain the role of **ribavirin** include subtle direct reduction of HCV replication, inhibition of host inosine monophosphate dehydrogenase activity (and associated depletion of guanosine pools), immune modulation, induction of virologic mutational catastrophe, and enhancement of IFN-stimulated gene expression. **Ribavirin**, despite its poorly understood mechanism of action, retains a modest role in supporting DAA agents as well (see below). IFN therapy results in activation of the JAK-STAT signal transduction pathway, which culminates in the intracellular elaboration of genes and their protein products that have antiviral properties. Hepatitis C proteins inhibit JAK-STAT signaling at several steps along the pathway, and exogenous IFN restores expression of IFN-stimulated genes and their antiviral effects.

Treatment with the combination of PEG IFN and **ribavirin** increased responsiveness (frequency of SVR) to as high as 55% overall—to $> 40\%$ in genotypes 1 and 4, and to $> 80\%$ in genotypes 2 and 3. Even in the absence of biochemical and virologic responses, histologic improvement occurred in approximately three-fourths of all treated patients. In chronic hepatitis C, ALT levels fall precipitously during therapy, and up to 90% of virologic responses are achieved within the first 12 weeks of therapy; responses thereafter are rare. Most relapses occur within the first 12 weeks after treatment; therefore, an SVR at week 12 posttreatment (SVR₁₂) is roughly equivalent to a 24-week SVR, and SVR₁₂ has become the new standard. SVRs are very durable; normal ALT, improved histology, and absence of HCV RNA in serum and liver have been documented a decade after successful therapy, and “relapses” 2 years after sustained responses are almost unheard of. Thus, an SVR to antiviral therapy of chronic hepatitis C is tantamount to a cure, which is followed by marked improvements in liver-disease outcomes (see below).

Patient variables that correlate with sustained virologic responsiveness to IFN-based therapy include favorable genotype (genotypes 2 and 3 as opposed to genotypes 1 and 4; genotype 1b as opposed to genotype 1a); low baseline HCV RNA level ($< 800,000$ IU/mL), low HCV quasispecies diversity, and histologically mild hepatitis and minimal fibrosis, especially absence of cirrhosis; immunocompetence, low liver iron levels, age < 40 ; female gender; and absence of obesity, insulin resistance, type 2 diabetes mellitus, and hepatic steatosis. High levels of HCV RNA, more histologically advanced liver disease, and high HCV quasispecies diversity all go hand in hand with advanced duration of infection and reduced IFN responsiveness. Also associated with poor responses to IFN-based therapy are African-American ethnicity (contributed to, but not explained entirely by, a higher proportion with genotype 1, slower early treatment viral kinetics, impaired HCV-specific immunity, and host genetic differences in *IL28B* alleles, described below), Latino ethnicity, and poor treatment adherence ($< 80\%$ of IFN and **ribavirin** doses and $< 80\%$ of prescribed duration of therapy). Ironically, patients whose

disease was least likely to progress were the ones *most* likely to respond to IFN and vice versa. For patients treated with combination IFN-ribavirin, therapy for those with genotype 1 usually required a full 48 weeks with SVRs in the range of 40–45%, whereas in those with genotypes 2 and 3, a 24-week course of therapy sufficed with SVRs in the range of 80% (although refined tailoring of treatment duration could be indicated based on rapidity of response or associated cofactors, see below).

Genetic changes in the virus may explain differences in treatment responsiveness in some patients (e.g., among patients with genotype 1b, responsiveness to IFN is enhanced in those with amino-acid-substitution mutations in the nonstructural protein 5A gene). As described above in the discussion of spontaneous recovery from acute hepatitis C, IFN gene variants discovered in genome-wide association studies were shown to have a substantial impact on responsiveness of patients with genotype 1 to antiviral therapy. In studies of patients treated with PEG IFN and ribavirin, variants of the *IL28B* SNP that code for IFN- λ 3 (a type III IFN, the receptors for which are more discretely distributed than IFN- α receptors and more concentrated in hepatocytes) correlate significantly with responsiveness. Patients homozygous for the C allele at this locus have the highest frequency of achieving an SVR (~80%), those homozygous for the T allele at this locus are least likely to achieve an SVR (~25%), and those heterozygous at this locus (C/T) have an intermediate level of responsiveness (SVRs in ~35%).

Side effects of IFN therapy are described in the section on treatment of chronic hepatitis B. The most pronounced side effect of ribavirin therapy is hemolysis—an expected reduction in hemoglobin of up to 2–3 g or in hematocrit of 5–10% but also a small, unpredictable proportion with profound, brisk hemolysis, resulting in symptomatic anemia; therefore, close monitoring of blood counts is crucial, and ribavirin should be avoided in patients with anemia or hemoglobinopathies; in patients with coronary artery disease or cerebrovascular disease, in whom anemia can precipitate an ischemic event; in patients with renal insufficiency (the drug is excreted renally); and in pregnancy (the drug is teratogenic, mandating scrupulous use of efficient contraception during, and for several months after, therapy in women of child-bearing age [because of their antiproliferative properties, IFNs also are contraindicated during pregnancy]). When symptomatic anemia occurs, ribavirin dose reductions or addition of erythropoietin to boost red blood cell levels may be required; erythropoietin was shown to improve patients' quality of life but not the likelihood of achieving an SVR. If ribavirin was stopped during therapy, SVR rates fell, but responsiveness could be maintained as long as ribavirin was not stopped and the total ribavirin dose exceeded 60% of the planned dose.

Ribavirin can also cause nasal and chest congestion, pruritus, and precipitation of gout. Combination IFN-ribavirin therapy is more difficult to tolerate than IFN monotherapy and more likely to lead to dose reductions and discontinuation of therapy.

Studies of viral kinetics have shown that despite a virion half-life in serum of only 2–3 h, the level of HCV is maintained by a high replication rate of 10^{12} hepatitis C virions per day. IFN- α blocks virion production or release with an efficacy that increases with increasing drug doses; moreover, the calculated death rate for infected cells during IFN therapy is inversely related to the level of HCV RNA. Patients with the most rapid death rate of infected hepatocytes are more likely to achieve undetectable HCV RNA at 3 months; in practice, failure to achieve an early virologic response (EVR), a ≥ 2 -log₁₀ reduction in HCV RNA by week 12, predicts failure to

experience a subsequent SVR. Similarly, patients in whom HCV RNA becomes undetectable within 4 weeks (i.e., who achieve a rapid virologic response [RVR]) have a very high likelihood of achieving an SVR (Fig. 334-2). Surprisingly, however, high-dose induction with IFN-based therapy did not yield higher SVR rates.

For the treatment of chronic hepatitis C, standard IFNs were supplanted beginning in 2001 by PEG IFNs. These have elimination times up to sevenfold longer than standard IFNs (i.e., a substantially longer half-life) and achieve prolonged concentrations, permitting administration once (rather than three times) a week. Instead of the frequent drug peaks (linked to side effects) and troughs (when drug is absent) associated with frequent administration of short-acting IFNs, administration of PEG IFNs results in drug concentrations that are more stable and sustained over time. Once-a-week PEG IFN monotherapy is twice as effective as monotherapy with its standard IFN counterpart, approaches the efficacy of combination standard IFN plus **ribavirin**, and is as well tolerated as standard IFNs, without more difficult-to-manage thrombocytopenia and leukopenia than standard IFNs. For most of the decade prior to 2011, when protease inhibitors were introduced for HCV genotype 1 (see below), the standard of care was a combination of PEG IFN plus **ribavirin** for all HCV genotypes.

Two PEG IFNs are available: PEG IFN- α 2b, a 12-kD, linear PEG molecule bound to IFN- α 2b, and PEG IFN- α 2a, a larger, 40-kD, branched PEG molecule bound to IFN- α 2a; because of its larger size and smaller volume of extravascular distribution, PEG IFN- α 2a can be given at a uniform dose independent of weight, whereas the dose of the smaller PEG IFN- α 2b, which has a much wider volume distribution, must be weight-based. The standard dose of PEG IFN α 2a was 180 μ g and of PEG IFN- α 2b 1.5 μ g/kg. The **ribavirin** dose adopted for both PEG IFNs was, for genotype 1, 1000 mg (for patients <75 kg) to 1200 mg (for patients \geq 75 kg) and, for genotypes 2 and 3, 800 mg; a broader **ribavirin** dose/weight range was approved subsequently for PEG IFN- α 2b in patients with genotype 1: <65 kg, 800 mg; 65–85 kg, 1000 mg; >85–105 kg, 1200 mg; and >105 kg, 1400 mg. For both drugs, recommended treatment durations were 48 weeks for genotype 1 and 24 weeks for genotypes 2 and 3 (somewhat more refractory, justifying a full 48 weeks especially for advanced hepatic fibrosis or cirrhosis and/or high-level HCV RNA). Between the two PEG IFNs, PEG IFN- α 2a appeared to be slightly better tolerated and slightly more effective than PEG IFN- α 2b in registration trials (SVRs for genotype 1: 41–51% vs 40–42%, respectively) as well as in subsequent head-to-head trials and a systematic review of randomized trials (SVR in genotypes 1–4: 48–55% vs 32–40%, respectively).

Until the 2011 introduction of protease inhibitors, unless **ribavirin** was contraindicated (see above), combination PEG IFN plus **ribavirin** was the recommended course of therapy. Even after the introduction of protease inhibitors for genotypes 1 and 4, however, PEG IFN-**ribavirin** remained the standard of care for patients with genotypes 2 and 3 until late 2013. For patients treated with combination PEG IFN-**ribavirin**, measurement of quantitative HCV RNA levels at 12 weeks was helpful in guiding therapy; if a 2- \log_{10} drop in HCV RNA had not been achieved by this time, chances for an SVR were negligible, and additional therapy was futile. If the 12-week HCV RNA had fallen by 2 \log_{10} (EVR), the chances for an SVR at the end of therapy were approximately two-thirds; if the 12-week HCV RNA was undetectable (“complete” EVR), the chances for an SVR exceeded 80% (Fig. 334-2).

The frequency of an SVR to PEG IFN-**ribavirin** therapy could be increased by tailoring therapy according to baseline variables and on-treatment virologic responsiveness. In patients with baseline variables weighing

against a response (e.g., HCV RNA >800,000 IU/mL, weight >85 kg), by raising the dose of PEG IFN (e.g., to as high as 270 µg of PEG IFN-α2a) and/or the dose of **ribavirin** to as high as 1600 mg daily (if tolerated or supplemented by erythropoietin); or by extending therapy from 48 to 72 weeks for patients with genotype 1 and a slow virologic response (i.e., failure of HCV RNA to fall rapidly to undetectable levels within 4 weeks [absence of a RVR]), SVR rates could be improved somewhat. In contradistinction, in patients with genotype 1 (and 4) who had a 4-week RVR (which occurred in ≤20%), especially in the subset with low baseline HCV RNA, abbreviating the duration of therapy to 24 weeks, resulted in SVR rates of ~90%. Responsiveness to IFN-ribavirin-based therapy was diminished in immunocompromised patients and in patients with HIV-HCV co-infection and contraindicated in patients with decompensated liver disease or end-stage renal disease. The cumbersome nature of IFN-ribavirin-based therapy (injections, complicated laboratory monitoring, side effects and poor tolerability, modest efficacy, variables and patient subsets associated with poor responsiveness, tailored therapy, futility rules, etc.) was supplanted eventually (in 2016) by DAAs for all genotypes (see below). Most of the variables associated with poor responsiveness to IFN-based therapy became irrelevant, and difficult-to-treat patient subpopulations began to experience responses to DAAs that were indistinguishable from responses in standard patients (see below).

Persons with chronic HCV infection have been shown to suffer increased liver-related mortality. On the other hand, successful antiviral therapy of chronic hepatitis C resulting in an SVR has been shown to improve survival (and to reduce the need for liver transplantation); to lower the risk of liver failure, liver-related death, and all-cause mortality; to slow the progression of chronic hepatitis C; and to reverse fibrosis and even cirrhosis. Whereas the 10-year and 20-year survival in the absence of an SVR is reduced in cirrhotic patients with chronic hepatitis C, survival at these intervals after an SVR has been found to be indistinguishable from that of the general population. Although successful treatment reduces mortality and liver failure (3-4-fold 10-year reduction) in cirrhotic patients (and in those with advanced fibrosis) and reduces the need for liver transplantation and the likelihood of HCC (14-fold 10-year reduction), the risk of liver-related death and HCC persists, albeit at a much reduced level, necessitating continued clinical monitoring and cancer surveillance after SVR in cirrhotics. On the other hand, in the absence of an SVR, IFN-based therapy does not reduce the risk of HCC. Similarly, for nonresponders to PEG IFN-ribavirin therapy, three trials of long-term maintenance therapy with PEG IFN showed no benefit in reducing the risk of histologic progression or clinical decompensation, including the development of HCC. Fortunately, PEG IFN-ribavirin nonresponders can now be retreated with DAAs and experience SVR rates comparable to those in treatment-naïve persons (see below).

FIRST-GENERATION PROTEASE INHIBITORS (2011–2013)

The HCV RNA genome encodes a single polyprotein, which is cleaved during and after translation by host and viral-encoded proteases. One protease involved in the cleavage of the viral polyprotein is an NS3/4A viral protein that has serine protease activity. Telaprevir and boceprevir are serine protease inhibitors that target NS3/4A. In 2011, telaprevir and boceprevir used in combination with PEG IFN and **ribavirin** were approved by the U.S. Food and Drug Administration (FDA) as the first oral DAA agents for the treatment of hepatitis C genotype 1 (not other genotypes) in adults with stable liver disease, both in patients who had not been treated before or who had failed previous treatment. Although now replaced by more effective, all-oral regimens, these first-in-class agents represented a breakthrough in the treatment of chronic hepatitis C and established milestones against which subsequent therapies could be measured.

Because resistance developed rapidly during monotherapy with telaprevir and boceprevir, these drugs had to be used in combination with PEG IFN and **ribavirin**. **Ribavirin** in particular appeared to reduce relapse rates significantly in protease inhibitor-based regimens, such that those who could not take or were intolerant to **ribavirin** were unlikely to benefit from the addition of these agents. Telaprevir and boceprevir regimens consisted of periods of triple therapy (protease inhibitor plus PEG IFN plus **ribavirin**) and periods of dual therapy (PEG IFN plus **ribavirin**). Telaprevir regimens began with 12 weeks of triple therapy followed by dual therapy of a duration based on HCV RNA status at weeks 4 and 12 (“response-guided therapy”) and prior treatment status. Boceprevir-based regimens consisted of a 4-week lead-in period of dual (PEG IFN–**ribavirin**) therapy followed by triple therapy and, in some instances, a further extension of dual therapy, with duration of response-guided therapy based on HCV RNA status at weeks 4, 8, and 24 and prior treatment status.

For patients with HCV genotype 1, protease inhibitors improved the frequency of RVRs and SVRs significantly as compared to PEG IFN plus **ribavirin** alone. In treatment-naïve patients, telaprevir-based SVRs were achieved in up to 79% of patients who received 12 weeks of triple therapy followed by 12–36 weeks of dual therapy, and among those with EVRs (undetectable HCV RNA at weeks 4 and 12) and response-guided therapy stopped at week 24 (12 weeks of triple therapy, then 12 weeks of dual therapy), SVRs occurred in 83–92%. In studies with boceprevir in treatment-naïve patients, SVRs occurred in 59–66% of patients, and among those with undetectable HCV RNA at 8 weeks, the SVR rate increased to 86–88%. Adding to the complexity of treatment with these protease inhibitors were absolute stopping rules for futility, that is, absence of HCV RNA reductions at critical treatment milestones, which were shown to be invariably predictive of nonresponse (telaprevir: HCV RNA >1000 IU/mL at weeks 4 or 12, or detectable at week 24; boceprevir: HCV RNA \geq 100 IU/mL at week 12, or detectable at week 24).

In patients previously treated unsuccessfully with PEG IFN plus **ribavirin**, telaprevir-based treatment achieved SVRs in 83–88% of prior relapsers, 54–59% of partial responders (HCV RNA reduced by $\geq 2 \log_{10}$ IU/mL but not to undetectable levels), and 29–33% of null responders (HCV RNA reduced by $< 2 \log_{10}$ IU/mL). With boceprevir, a similar degradation in SVR rate occurred as a function of prior responsiveness—in 75% of prior relapsers, in 40–52% of previous partial responders; in ~30–40% of null responders. In a substantial proportion of protease inhibitor nonresponders, resistance-associated substitutions (RASs, previously referred to as resistance-associated variants, RAVs) could be identified, but these variants were not archived, and wild-type HCV reemerged in almost all cases within 1.5 to 2 years. SVRs to these protease inhibitors were highest in prior relapsers and treatment-naïve patients (white > black ethnicity), lower in prior partial responders, lower still in prior null responders, and lowest in cirrhotic prior null responders, for whom no benefit accrued over PEG IFN/**ribavirin** treatment. Responses to protease inhibitor triple-drug regimens were higher in patients with *IL28BC* than non-C genotypes, HCV genotype 1b than genotype 1a, less advanced than more advanced fibrosis stage, whites than blacks, lower body mass index (BMI) than elevated BMI, and, for boceprevir, achievement of a $> 1 \log_{10}$ HCV RNA reduction during 4 weeks of PEG IFN–**ribavirin** lead-in therapy. Age and HCV RNA level were less influential and insulin resistance was noninfluential on response to these antiviral agents.

Both of these protease inhibitors had substantial toxicities. Telaprevir was associated with a severe, generalized (trunk and extremities), often confluent, maculopapular, pruritic rash in ~6% of treated patients

(that required careful dermatologic monitoring in all patients and systemic corticosteroid therapy in the most severely affected). Other common side effects included pruritus, rectal burning, nausea, diarrhea, fatigue, dysgeusia (altered or unpleasant taste), and anemia, which required close monitoring, could be relatively refractory, occasionally requiring transfusion and even hospitalization (especially in cirrhotic prior nonresponders). Anemia occurred in half of boceprevir-treated patients, neutropenia in up to 30% and thrombocytopenia in 3–4%. Other side effects of boceprevir include fatigue, nausea, headache, dysgeusia, dry mouth, vomiting, and diarrhea.

Both drugs came with an inconveniently high pill burden and had to be administered every 8 hours with food (TVR with a 20-g fat meal). Use of protease inhibitors was further complicated by numerous drug-drug interactions. As telaprevir and boceprevir are both eliminated by and inhibit CYP3A4, these agents could not be administered with other medications that induce CYP3A4 or are dependent on CYP3A4 for elimination. Care had to be taken to examine for any potential interactions between these protease inhibitors and other medications the patient was taking, and a convenient website became available to check for such drug-drug interactions (www.hep-druginteractions.org).

Despite the improvement in SVRs with protease-inhibitor-based regimens for genotype 1 compared to PEG IFN-ribavirin (e.g., in treatment-naïve patients 66–79% vs 38–44%), triple-drug protease-inhibitor therapy was hampered by amplified intolerability, the complexity of response-guided regimens and futility stopping rules, the inconvenience of thrice-daily dosing with meals and a high pill burden, the need for PEG IFN injections and ribavirin with all their intolerability, and multiple drug-drug interactions. Moreover, side effects appeared to be more severe and burdensome once these drugs entered practice, especially in cirrhotic nonresponders, in whom studies reported from Europe showed serious adverse events in up to 45% and deaths in up to 3%. All these issues, as well as rapidly accelerating progress on next-generation and all-oral DAA therapy (see below), conspired to temper enthusiasm for these new antivirals; after a brief stint as recommended therapy (2011–2013), these drugs became obsolete and are no longer recommended.

CONTEMPORARY DIRECT-ACTING ANTIVIRAL COMBINATION THERAPY (2013–)

Since late 2013, the number of new antiviral agents for hepatitis C has expanded substantially, and, currently, PEG IFN-based treatments have been supplanted by six therapeutic regimens: all oral, IFN-free, highly efficacious (>95% SVR), well tolerated, with high barriers to resistance, simple dosing and low pill burdens, treatment durations as brief as 8 to 12 weeks, and, in many cases, pangenotypic efficacy (**Table 334-6**). These drugs are distributed among three classes of DAAs: NS3/4 protease inhibitors (which cleave the single HCV polyprotein into constituent structural and nonstructural proteins), NS5B nucleoside and nonnucleoside polymerase inhibitors (which interfere with the RNA-dependent RNA polymerase [a replicase] involved in synthesis of viral RNA), and NS5A inhibitors (which interfere with a membrane-associated phosphoprotein essential to the HVC RNA replication complex).

The first of the new DAA agents (approved in November 2013) was simeprevir, a second-generation protease inhibitor for genotype 1, followed shortly thereafter (December 2013) by sofosbuvir, a pangenotypic nucleoside polymerase inhibitor. For genotype 1, both of these agents had to be combined with PEG IFN and ribavirin; for genotypes 2 and 3, sofosbuvir was administered with ribavirin, without PEG IFN; however, these treatment

regimens have been supplanted by combinations of all-oral, IFN-free, DAAs, and **ribavirin** is rarely needed, retained only for very limited indications.

Simeprevir: When simeprevir was used with PEG IFN, its efficacy (genotype 1b > 1a) was similar to that of first-generation protease inhibitors, but required only once-a-day dosing without the complexity of response-guided therapy. Similar to first-generation protease inhibitors, simeprevir was hampered by many drug-drug interactions and side effects (including photosensitivity, rash, and mild hyperbilirubinemia); moreover, patients, with HCV NS3 polymorphism Q80K had markedly reduced drug efficacy, necessitating pretreatment genetic testing and disqualifying a substantial proportion (approximately a third) of potential treatment candidates. Little about simeprevir supported its adoption in combination with PEG IFN and **ribavirin**. On the other hand, the combination of simeprevir (150 mg) along with **sofosbuvir** (400 mg) for 12 weeks was found to be effective in treatment-naïve (97% SVR₁₂) or treatment-experienced (95% SVR₁₂) patients without cirrhosis and in treatment-naïve (88% SVR₁₂) or treatment-refractory (79% SVR₁₂) patients with cirrhosis (it remains one of the recommended regimens for genotype 1).

Sofosbuvir: **Sofosbuvir**, the first nonprotease inhibitor DAA to be approved, has an excellent profile—high potency, high barrier to resistance, pangenotypic activity, very well tolerated with limited adverse effects (most commonly mild fatigue, insomnia, headache, and nausea), once-daily oral administration, and relative freedom from major drug-drug interactions. **Sofosbuvir** has efficacy in all genotypes (1 to 6); in treatment-naïve subjects and prior nonresponders to PEG IFN-based and protease-inhibitor-based therapy; with PEG IFN-RBV or in IFN-free regimens; in combination with RBV or with NS5A inhibitors; and for treatment periods as brief as 8 to 12 weeks to as long as 24 weeks. Currently, **sofosbuvir** is used in combination with either the protease inhibitor simeprevir (as described above) or, more commonly, with one of three NS5A inhibitors. Thus, **sofosbuvir** is a component of four of the six recommended DAA regimens for genotype 1, two of the four regimens for genotype 4, and both of the regimens for genotypes 2, 3, 5, and 6 ([Table 334-6](#)).

Sofosbuvir/ledipasvir: The DAA combination that has had a dominant role in the treatment of hepatitis C is **sofosbuvir** (400 mg) plus the NS5A inhibitor ledipasvir (90 mg) in a once-a-day, fixed-dose, single pill, approved in October 2014 for genotype 1 and in November 2015 for genotypes 4, 5, and 6. Phase-III trials were conducted in treatment-naïve noncirrhotic patients, in treatment-naïve cirrhotic and noncirrhotic patients, and in treatment-experienced cirrhotic and noncirrhotic patients treated for 8, 12, or 24 weeks, both with and without **ribavirin**. In treatment-naïve noncirrhotics, an SVR₁₂ was achieved in 97–99% of subjects, and no benefit was observed by extending therapy from 12 to 24 weeks or by adding **ribavirin**. Moreover, for treatment-naïve, noncirrhotic patients with baseline HCV RNA <6 × 10⁶ IU/mL, a treatment duration of 8 weeks was as effective as one of 12 weeks (94–95% SVR₁₂), which may be a consideration for a proportion of patients. In cirrhotic patients, SVR₁₂ was achieved in 97–100% of treatment-naïve subjects (no advantage of extending therapy from 12 to 24 weeks or of adding **ribavirin**); however, for cirrhotic prior nonresponders to IFN-based therapy, 12 weeks of therapy was inferior (86% SVR₁₂) to 24 weeks of therapy (100% SVR₁₂). This combination, which is equally effective in patients with HIV-HCV co-infection and in African-American patients, has been shown to be highly effective in patients with decompensated cirrhosis and in patients with hepatitis C after liver

transplantation and after kidney transplantation. On the other hand, the safety and efficacy of **sofosbuvir**/ledipasvir in patients with advanced renal failure have not been established, and all sofosbuvir-containing regimens can be associated with severe bradycardia in patients taking the antiarrhythmic agent amiodarone, especially along with beta blockers; sofosbuvir-containing combinations are contraindicated with amiodarone. Drug-drug interactions are few, but P-gp inducers, like St. John's wort and rifampin, and proton-pump gastric acid inhibitors, like **omeprazole**, may reduce **sofosbuvir**/ledipasvir concentrations. Generally, responsiveness to **sofosbuvir**/ledipasvir is not reduced in patients with baseline RASs to these agents, with the exception of treatment-experienced patients who have baseline NS5A RASs (for whom EASL recommends adding **ribavirin** or, if **ribavirin** is contraindicated, extending treatment to 24 weeks).

Paritaprevir/ritonavir, ombitasvir, and dasabuvir: The combination of ritonavir (100 mg)-boosted paritaprevir (150 mg), a protease inhibitor; ombitasvir (25 mg), an NS5A inhibitor; dasabuvir (250 mg), a nonnucleoside polymerase inhibitor; ± weight-based **ribavirin** (total of five drugs) was approved in December 2014 for genotypes 1 and 4. Paritaprevir/ritonavir and ombitasvir, formulated in a single tablet, are taken once daily, and both dasabuvir (a separate pill) and weight-based **ribavirin** (when included in the regimen) are taken twice daily. In clinical trials, this combination achieved SVR₁₂ rates of 87–100% in treatment-naïve and treatment-experienced patients with genotype 1; without **ribavirin**, this combination in genotype 1a is ~7% less responsive than genotype 1b. Therefore, in treatment-naïve patients with genotype 1a, this combination is administered *with* **ribavirin** for 12 weeks in the absence of cirrhosis (95–97% SVR₁₂) or for 24 weeks in the presence of compensated cirrhosis (94% SVR₁₂), while in patients with genotype 1b, the combination does not require **ribavirin**, and the duration of therapy is 12 weeks for both noncirrhotics and cirrhotics (99–100% SVR₁₂). In prior nonresponders without cirrhosis, the combination is administered for 12 weeks, *with* **ribavirin** in genotype 1a (96% SVR₁₂), *without* **ribavirin** in genotype 1b (100% SVR₁₂). In prior nonresponders with cirrhosis, the combination is administered for 24 weeks *with* **ribavirin** in genotype 1a (SVR₁₂ 100% in prior relapsers and partial responders, 95% in prior null responders [in whom treatment without **ribavirin** was associated with an 80% SVR₁₂]), but only for 12 weeks and *without* **ribavirin** in genotype 1b (100% SVR₁₂). For genotype 4, the regimen is given for 12 weeks with **ribavirin**, but without dasabuvir in treatment-naïve and treatment-experienced patients (100% SVR₁₂), including those with compensated cirrhosis. In July 2016, the FDA approved a long-acting formulation of dasabuvir, allowing once-a-day instead of twice-a-day treatment; for genotype 1a, twice-daily **ribavirin** dosing remains.

This combination is well tolerated with generally mild side effects, for example, fatigue, asthenia, insomnia, headache, and pruritus. Hyperbilirubinemia (primarily unconjugated) and elevations in alanine aminotransferase activity may occur but resolve during or shortly after treatment. Because of occasional hyperbilirubinemia and potential hepatotoxicity (FDA warning letter issued October 2015 regarding hepatic failure/decompensation reported in treated cirrhotic patients), this combination is not recommended in patients with decompensated cirrhosis, and treated cirrhotic patients should be monitored closely for decompensation; however, the safety and efficacy of this combination have been demonstrated for patients with advanced renal insufficiency. Similar to other regimens containing protease inhibitors, drug-drug interactions are common with other drugs that induce CYP3A4 or are dependent on CYP3A4 for elimination.

Checking for potential drug-drug interactions is important prior to initiating therapy with this drug combination (www.hep-druginteractions.org). Responsiveness to this multidrug regimen is not reduced in patients with baseline RASs to these agents.

Compared to [sofosbuvir](#)/ledipasvir, this regimen has the disadvantage of requiring twice-a-day [ribavirin](#) therapy for genotype 1a and of being contraindicated in decompensated cirrhosis; however, it has the advantage of offering a 12-week, ribavirin-free regimen for prior null responders with cirrhosis and providing an option for patients with renal failure.

[Sofosbuvir](#) and [Daclatasvir](#): Daclatasvir, an NS5A inhibitor, along with the polymerase inhibitor [sofosbuvir](#), was approved by the FDA in July 2015 for genotype 3 and in February 2016 for genotype 1 (AASLD-Infectious Diseases Society of America [IDSA] guidelines [see below] include its recommendation as well for genotype 2; in August 2014, this combination was approved in Europe for genotypes 1, 2, 3, and 4, and EASL recommends it for all these genotypes as well as for genotypes 5 and 6). At the time of its approval for genotype 3, daclatasvir filled a need inadequately met by other available combination DAAs. Although data on genotype 3 are the most robust, clinical trials of this combination in genotypes 1 and 2 support its efficacy and recommendations for first-line (genotype 1) and alternative (genotype 2) treatment, in some cases with [ribavirin](#) (Table 334-6). Daclatasvir, a 60-mg tablet, and [sofosbuvir](#), a separate 400 mg tablet are taken once-a-day for 12 to 24 weeks.

In clinical trials among treatment-naïve or treatment-experienced patients, SVR₁₂ rates for 12 weeks of daclatasvir plus [sofosbuvir](#) were 98% with genotype 1 (comparable results in genotypes 1a and 1b), 92% for genotype 2, and 89% for genotype 3. For noncirrhotic patients, the addition of [ribavirin](#) or the extension of therapy to 24 weeks did not improve efficacy. In patients with compensated cirrhosis, limited prospective data and data from observational cohorts suggested that extending therapy to 24 weeks, with or without [ribavirin](#), improved efficacy. In cirrhotics, SVR₁₂ was achieved in 93% with Child Class-Pugh A and B but in only 56% with Class-C decompensated cirrhosis. For patients with genotype 3 and cirrhosis, the combination was effective in treatment-naïve patients (94% SVR₁₂), but less so in prior nonresponders (69% SVR₁₂). Outcomes in patients with HIV-HCV co-infection were comparable.

Like other sofosbuvir-NS5A inhibitor combinations, daclatasvir plus [sofosbuvir](#) is well tolerated (mild fatigue, headache, nausea, diarrhea in 5–14%), but can cause severe bradycardia when administered with amiodarone (contraindicated), especially along with beta blockers. Because daclatasvir is a substrate for CYP3A, CYP3A inducers can reduce daclatasvir levels, and CYP3A inhibitors reduce daclatasvir levels. Similarly, daclatasvir, an inhibitor of P-gp, OATP1B1 and 1B3, and BCP, can increase the levels of drugs that are substrates of these transporters. As noted above for other DAAs, checking for potential drug-drug interactions is advisable prior to initiating therapy (www.hep-druginteractions.org). Responsiveness to daclatasvir-containing drug-combination therapy is reduced in cirrhotic patients with genotype 1a and in both cirrhotic and noncirrhotic patients with genotype 3 who have baseline daclatasvir-associated NS5A RASs.

Although daclatasvir-sofosbuvir is approved for genotypes 1 and 3 and recommended as an alternative for genotype 2, better documented efficacy and simplicity of other regimens have limited the popularity of this drug combination.

Elbasvir/Grazoprevir: Elbasvir (50 mg), an NS5A inhibitor, combined in a single, fixed-dose pill with grazoprevir (100 mg), an NS3/4 protease inhibitor, was approved in January 2016 as a once-a-day (with or without food) treatment for genotypes 1 and 4. In clinical trials, a 12-week course was effective in treatment-naïve and treatment-experienced patients without cirrhosis or with compensated cirrhosis. In treatment-naïve patients, this combination yielded an SVR₁₂ in 92% of patients with genotype 1a, 99% with genotype 1b, and 100% with genotype 4 (very small numbers, however); 10 patients with genotype 6 were included, but only 80% achieved SVR₁₂. Cirrhotic and noncirrhotic patients had comparable rates of SVR₁₂, 97% and 94%, respectively. For this drug combination, however, ~11% of patients with genotype 1a harbor NS5A polymorphisms, that is, RASs, at baseline. If present, these NS5A RASs reduce efficacy of elbasvir/grazoprevir (unlike baseline RASs to the most of the other combination DAA regimens described above and below) from 99% to 58% in treatment-naïve patients. Therefore, all patients with genotype 1a require baseline RAS testing; if these RASs are present, treatment extension to 16 weeks and the addition of weight-based [ribavirin](#) bring the SVR₁₂ up to expected levels of close to 100%. In treatment-experienced patients, both extending treatment to 16 weeks and adding [ribavirin](#) were studied; however, generally, in the absence of baseline NS5A RASs, SVR₁₂ rates were not increased over those without [ribavirin](#) for 12 weeks (94–97%). For genotype 1a, among prior nonresponders to PEG IFN/[ribavirin](#), 12 weeks of elbasvir/grazoprevir suffices without [ribavirin](#) except for patients with baseline NS5A RASs, who require 16 weeks of therapy and [ribavirin](#). Among nonresponders to prior protease-inhibitor therapy, even in the absence of baseline NS5A RASs, [ribavirin](#) should be added to a 12-week regimen; in the presence of baseline NS5A RASs, treatment should be extended to 16 weeks and [ribavirin](#) added. For genotype 1b, NS5A RASs are not an issue, and the only subgroup requiring modification of a 12-week course of therapy are prior nonresponders to protease-inhibitor regimens, for whom [ribavirin](#) is added. For genotype 4, the recommended regimen for all prior nonresponders (whether to PEG IFN/[ribavirin](#) or protease inhibitor regimens) is 16 weeks of elbasvir/grazoprevir plus [ribavirin](#) (Table 334-6).

This combination is just as effective in patients with HIV-HCV co-infection and in patients with advanced renal failure (including those requiring hemodialysis); however, it is contraindicated in decompensated cirrhosis. Like other protease inhibitor regimens, elbasvir/grazoprevir can be associated with aminotransferase elevations and potential hepatotoxicity; because these drugs are excreted by the liver, in decompensated liver disease, plasma drug concentrations may become elevated substantially. Therefore, all treated patients should have alanine aminotransferase screening periodically during therapy, and the drug should be stopped for elevations exceeding 10-fold or for elevations of conjugated bilirubin, alkaline phosphatase, or prothrombin time.

Elbasvir/grazoprevir is well tolerated, with only low levels of mild adverse effects (fatigue, headache, nausea in 5–11%) seen just as frequently in placebo recipients. Both elbasvir and grazoprevir are substrates for CYP3A and are subject to multiple potential drug-drug interactions. Therefore, this combination should not be used with potent CYP3A inducers; conversely, CYP3A and OATP1B1 inhibitors can lead to untoward elevations of plasma elbasvir/grazoprevir concentrations. Checking for potential drug-drug interactions is advisable prior to initiating therapy (www.hep-druginteractions.org).

Compared to other available regimens for genotypes 1 and 4, elbasvir/grazoprevir has the disadvantage/inconvenience of requiring baseline NS5A RAS testing but the advantages of a comparable regimen for cirrhotics and noncirrhotics, for treatment-naïve and treatment-experienced patients, and for patients with normal renal function and with renal failure.

Sofosbuvir/velpatasvir: The combination in a single, fixed-dose pill of velpatasvir (100 mg), a highly potent, pangenotypic NS5A inhibitor, along with the polymerase inhibitor **sofosbuvir** (400 mg) was approved in June 2016 for genotypes 1–6, in treatment-naïve and treatment-experienced noncirrhotics and cirrhotics. **Ribavirin** is not required, including in patients with genotypes 2 and 3, except in patients with decompensated cirrhosis.

In a series of clinical trials, this combination for 12 weeks in the absence of **ribavirin** was shown to yield 99% SVR₁₂ (range 97–100%) in genotypes 1, 2, 4, 5, and 6 and 95% in genotype 3. Baseline NS5A RASs had no impact on responsiveness.

Prior to the availability of this drug combination, patients with genotype 3, especially those with cirrhosis and prior null response to other therapies, proved to be the most refractory subset of patients. In treatment-naïve patients with genotype 3, 12 weeks of **sofosbuvir**/velpatasvir (95% SVR₁₂) was superior to 24 weeks of **sofosbuvir** plus **ribavirin** (80% SVR₁₂). In patients with genotype 3, the combination of **sofosbuvir**/velpatasvir for 12 weeks was comparable in noncirrhotics (97% SVR₁₂) and cirrhotics (91% SVR₁₂) and in treatment-naïve (97% SVR₁₂) and treatment-experienced (90% SVR₁₂) patients, superior in all these categories to 24 weeks of **sofosbuvir** plus **ribavirin** (87%, 66%, 86%, and 63%, respectively). In cirrhotic null responders, most available IFN-free regimens for genotype 3 (including daclatasvir plus **sofosbuvir**, approved specifically for this genotype) achieved SVR₁₂ rates in the range of ~60–75%, while the combination of PEG IFN, **ribavirin**, and **sofosbuvir** could boost SVR₁₂ to the mid-80% range. For treatment-experienced patients with genotype 3, **sofosbuvir**/velpatasvir in noncirrhotics and cirrhotics had similarly high efficacy (91% and 89% SVR₁₂, respectively); this was the highest recorded SVR₁₂ for genotype-3 cirrhotic null responders treated with IFN-free DAA regimens. Finally, in patients with genotypes 1–4 and 6 and with decompensated, Class-B cirrhosis (55% treatment-experienced), **sofosbuvir**/velpatasvir plus **ribavirin** for 12 weeks yielded an SVR₁₂ in 94%; this result was better than **sofosbuvir**/velpatasvir without **ribavirin** for 12 weeks (83% SVR₁₂) or 24 weeks (86% SVR₁₂).

Like other all-oral DAAs, **sofosbuvir**/velpatasvir was very well tolerated; in noncirrhotic and compensated cirrhotic patients, mild headache and fatigue was seen in >10%—this occurred in a comparable proportion of placebo recipients; in decompensated cirrhosis, mild fatigue, headache, nausea, insomnia, diarrhea, and anemia (**ribavirin** was part of the regimen) was seen in >10%. Like other sofosbuvir-containing regimens, **sofosbuvir**/velpatasvir should not be administered along with amiodarone (potential serious bradycardia); in addition, P-gp inducers and moderate-to-potent CYP3A inducers can reduce plasma levels of **sofosbuvir** and/or velpatasvir. Checking for drug-drug interactions prior to therapy is advisable (www.hep-druginteractions.org). Baseline RASs do not influence responsiveness to this combination.

FUTURE DIRECT-ACTING ANTIVIRAL COMBINATION THERAPY (2017–)

Most treatment needs have been met by contemporary DAA regimens described above; however, several additional, highly potent, pangenotypic drug combinations are in development. For example, an investigative protease inhibitor (voxilaprevir) added to the polymerase inhibitor/NS5A inhibitor combination of [sofosbuvir](#)/velpatasvir yields a very well tolerated *triple-drug* combination with 97% SVR₁₂ across all HCV genotypes and patient subgroups. These include noncirrhotic/cirrhotic, treatment-naïve/treatment-experienced groups, including those who had prior NS5A treatment and results were independent of the number of prior DAA drug classes received; no effects of baseline NS5A RASs were noted. Several experimental combinations may allow even briefer durations of therapy. In a small, exploratory trial, a *6-week* combination of [sofosbuvir](#) plus an experimental pangenotypic, very high potency, very low resistance NS5A inhibitor (odalasvir) achieved SVR₁₂ in 100% of 12 patients with genotype 1. Similarly, in a *6-week* triple combination of odalasvir with the protease inhibitor simeprevir and an experimental polymerase inhibitor (“AL-335”), SVR₁₂ was observed in 100% of 20 treatment-naïve noncirrhotic patients with genotype 1. In phase-II clinical trials, *8 weeks* of an experimental combination of two high-potency, pangenotypic DAAs, a protease inhibitor (“ABT-493”) plus an NS5A inhibitor (“ABT-530”), yielded 100% SVR₁₂ in treatment-naïve noncirrhotic patients with genotypes 1, 2, and 3. In cirrhotics with genotype 3 and in patients with genotypes 4, 5, and 6, 12 weeks of therapy with this DAA combination yielded 100% SVR₁₂. In patients with prior DAA treatment failure, 12 weeks of this double-combination sufficed to achieve a $\geq 95\%$ SVR₁₂; neither baseline NS5A nor protease inhibitor RASs influenced SVR₁₂ rates. No safety issues have been encountered, and the potential for drug-drug interactions is limited. These promising combinations are undergoing phase-II and phase-III trials.

Less advanced is the development of inhibitors of host proteins, such as oral, nonimmunosuppressive inhibitors of cyclophilin A (which interacts with NS5A during HCV replication) and subcutaneous antisense antagonists of host liver-expressed micro-RNA-122 (which promotes HCV replication). Given the accelerated progress of all-oral, short-treatment-duration, high-efficacy, DAAs, these alternative approaches may not be practical or competitive; moreover, development of both approaches has been retarded by emerging toxicities such as pancreatitis associated with cyclophilin inhibitors and jaundice associated with micro-RNA-122.

Although data on the impact of DAAs on the natural history of chronic hepatitis C are still limited, preliminary findings are that successful therapy is associated with a gradual reduction in fibrosis progression and a regression of advanced fibrosis (cirrhosis), improvement in survival among patients with decompensated cirrhosis, and a decline in the number of patients with hepatitis C being referred for liver transplantation. Based on the known prevalence, natural history, and rate of progression of chronic hepatitis C and on the efficacy of DAA therapies and their impact on the complications of hepatitis C, modeling estimates have suggested that the availability and application of these therapies have the potential to reduce the hepatitis C-associated disease burden including liver-related death, HCC, decompensated cirrhosis, and liver transplantation by 50–70% between 2015 and 2050.

TREATMENT RECOMMENDATIONS

Because the pace of new drug development and approval has been so rapid, the AASLD and the IDSA have been providing a consensus of updated treatment recommendations for patients with hepatitis C; these recommendations, which continue to be revised regularly based on new data, are available online at www.hcvguidelines.org and should be consulted before initiating therapy (Table 334-6). The EASL issues similar (but not identical) treatment recommendations annually for hepatitis C (www.easl.eu), most recently in September 2016. Divergences between AASLD-IDSA and EASL recommendations are noted in Table 334-6.

Prior to therapy, HCV genotype should be determined, because the genotype dictates which treatment regimens are indicated (Table 334-6). Monitoring of serum HCV RNA levels pretreatment, during treatment, and posttreatment is crucial in assessing response to therapy; moreover, the baseline level may contribute to determining the duration of therapy (e.g., in noncirrhotic patients with genotype 1 and HCV RNA $<6 \times 10^6$ IU/mL, 8 [instead of the usual 12] weeks of sofosbuvir/ledipasvir may be a consideration). The goal of treatment is to eradicate HCV RNA during therapy and to document that the virus remains undetectable for at least 12 weeks after completion of therapy (SVR₁₂). Several reports have appeared describing hepatitis B reactivation, often severe, during and after DAA therapy in patients coinfecting with HCV and HBV who were not being treated for their HBV infections. Therefore, screening for HBV infection is recommended prior to initiating DAA therapy for hepatitis C (which should have been done to determine HBV-immunity status as a prelude to recommended hepatitis B vaccination in patients with chronic hepatitis C), and therapy for HBV infection (for those meeting HBV treatment criteria, see above) should be initiated prior to or simultaneously with HCV therapy.

INDICATIONS FOR ANTIVIRAL THERAPY

Patients with chronic hepatitis C who have detectable HCV RNA in serum, whether or not aminotransferase levels are increased, and chronic hepatitis of any grade and stage are candidates for antiviral therapy with DAA agents. The only exception would be patients with short life expectancies, for whom treating hepatitis C would have no influence on longevity. Certainly, for patients with advanced liver disease, early treatment merits a high priority. Although patients with persistently normal aminotransferase activity tend to progress histologically very slowly or not at all, they respond to antiviral therapy just as well as do patients with elevated aminotransferase levels; therefore, such patients are potential candidates for antiviral therapy. As noted above, antiviral therapy has been shown to improve survival and complication-free survival and to slow progression of and to reverse fibrosis.

HCV genotype determines the regimen to be selected (Table 334-6). Similarly, the absence or presence of cirrhosis/advanced fibrosis determines the treatment options from which to select, including the antiviral agents to be used, the duration of therapy, and the need for ribavirin (Table 334-6). A pretreatment liver biopsy to assess histologic grade and stage provides substantial information about progression of hepatitis C in the past, has prognostic value for future progression, and can identify such histologic factors as steatosis and stage of fibrosis, which can influence responsiveness to therapy. As therapy has improved for patients with a broad range of histologic severity, and as noninvasive measures of the stage of fibrosis (e.g., assessment of liver elasticity by imaging) have gained in accuracy and popularity, noninvasive approaches have supplanted histology in most cases. If cirrhosis/advanced fibrosis is present prior to therapy, the risk of HCC, although reduced substantially by successful therapy, is not eliminated, and twice yearly posttreatment imaging for HCC

surveillance (and endoscopic surveillance for esophageal varices at intervals of 1–3 years) is indicated even after an SVR. In patients with low-level fibrosis at baseline, achievement of an SVR allows the cessation of such surveillance.

Patients who have relapsed after, or failed to respond to, a course of IFN-based or DAA agent-based therapy are candidates for retreatment with a DAA therapy regimen (Table 334-6). For patients who have failed to respond to a DAA combination, options include increasing the duration of therapy with the failed regimen, adding [ribavirin](#), or changing the drug class (e.g., after failed protease and polymerase inhibitors, switching to an NS5A-containing combination). In the presence of cirrhosis or a need for urgent retreatment, patients who have failed protease inhibitor plus polymerase inhibitor combination therapy or who have failed an NS5A combination are candidates for RAS testing and tailored therapy based on such resistance testing. If reliable RAS testing is not available, adding [ribavirin](#) or extending the duration of therapy are options. For prior nonresponders to IFN-based therapy, NS5A inhibitor-containing regimens are highly effective; however, reduced responsiveness can be encountered, especially in cirrhotic patients. For this relatively refractory group, ideally, the most potent/effective NS5A regimen should be selected to give such patients the best chance of responding and to avoid treatment-emergent NS5A RASs. Additional details for treatment of such patient subgroups can be found at www.hcvguidelines.org.

Persons with acute hepatitis C are also candidates for antiviral therapy (Chap. 332) with the same DAA agents approved for chronic hepatitis C; delaying the initiation of therapy for an observation period of 12–16 weeks (and even up to 6 months) has been recommended to allow for spontaneous recovery, especially in light of the fact that most cases of acute hepatitis C are not clinically severe or rapidly progressive. The duration of therapy for acute hepatitis C has not been determined definitively; however, in a small study of 20 patients, 6 weeks of [sofosbuvir](#)/ledipasvir sufficed for a 100% SVR₁₂. According to 2016 EASL recommendations, patients with acute hepatitis C should be treated for 8 weeks with a genotype-appropriate DAA regimen consisting of [sofosbuvir](#) plus one of the three approved NS5A inhibitors without [ribavirin](#) (extended to 12 weeks for patients with acute hepatitis C and HIV co-infection or for patients with acute hepatitis C and a baseline HCV RNA level >1 million IU/mL). In patients with biochemically and histologically mild chronic hepatitis C, the rate of progression is slow; however, such patients respond just as well to antiviral therapy as those with elevated aminotransferase levels and more histologically severe hepatitis. Because of the high cost of DAA treatments, initially a higher priority was assigned to patients with advanced fibrosis/cirrhosis; however, this controversial approach was relied upon by some medical insurers and pharmacy benefit management organizations to withhold therapy from patients with low-level fibrosis. Unfortunately, delaying therapy until fibrosis becomes advanced misses the opportunity to prevent all the dire consequences of chronic hepatitis C (liver failure, death/transplantation, HCC), which can be reduced, but not eliminated completely once advanced fibrosis is established. Therefore, therapy for patients with mild disease is justified as well as cost-effective.

Patients with compensated cirrhosis can respond to therapy, and their likelihood of a sustained response with DAAs is comparable to that in noncirrhotics. Patients with decompensated cirrhosis, who were not candidates for IFN-based antiviral therapy, respond well to DAA therapy regimens consisting of combinations of polymerase inhibitors and NS5A inhibitors (e.g., [sofosbuvir](#)/ledipasvir, [sofosbuvir](#)/velpatasvir); however, protease-inhibitor-containing combinations have been associated with potential hepatotoxicity and hepatic

decompensation and are contraindicated in this patient subset. Patients with decompensated cirrhosis should be referred to a liver transplantation center. DAAs are highly effective not only for patients with end-stage liver disease awaiting liver transplantation but also for patients with recurrent hepatitis C after liver transplantation. Ideally, patients should be treated prior to liver transplantation; however, a concern is that eradication of HCV infection will disqualify such patients from accepting donor livers from persons with HCV infection, thus contracting the potential donor pool and limiting accessibility to donor organs and timely transplantation. In addition, responsiveness to DAA therapy appears to be reduced in patients with decompensated cirrhosis and with high model for end-stage liver disease (MELD) scores; in this subgroup, responsiveness after liver transplantation would be substantially better. Therefore, advocacy has been expressed (recommended by EASL) for postponing DAA therapy in patients with high-MELD HCV-associated end-stage liver disease until after liver transplantation; the decision whether to treat pretransplantation or posttransplantation should be individualized thoughtfully for each patient, based on such factors as MELD score, time anticipated prior to availability of a donor organ, relative clinical stability, and co-morbidities ([Chap. 338](#)). The cutaneous and renal vasculitis of HCV-associated essential mixed cryoglobulinemia ([Chap. 332](#)) may respond to antiviral therapy, but sustained responses were rare after discontinuation of therapy in the IFN era, and prolonged, potentially indefinite, therapy was recommended. Now that more effective DAAs are available, a 12-week course of sofosbuvir-based combination therapy has been shown to yield an SVR₁₂ rate exceeding 80% in cryoglobulinemic vasculitis. Anecdotal reports suggest that IFN-based antiviral therapy may be effective in porphyria cutanea tarda or lichen planus associated with hepatitis C; whether the more appealing DAAs are effective in these groups remains to be documented.

In patients with HCV/HIV co-infection, hepatitis C is more progressive and severe than in HCV-monoinfected patients. Although patients with HCV/HIV co-infection responded less well to IFN-based antiviral therapy for hepatitis C, they respond as well as patients with HCV infection alone to DAA combination regimens. In HCV/HIV-infected patients, [ribavirin](#) can potentiate the toxicity of didanosine (e.g., lactic acidosis) and the lipodystrophy of stavudine, and zidovudine can exacerbate ribavirin-associated hemolytic anemia; therefore, these drug combinations should be avoided.

Patients with a history of injection drug use and alcoholism can be treated successfully for chronic hepatitis C, preferably in conjunction with drug and alcohol treatment programs. Moreover, because injection-drug users, as a source of transmission to others, account disproportionately for perpetuating the spread of HCV infection in the population, the impact of treating active injection-drug users is amplified by reducing such transmission. The approved oral combinations of DAAs are effective in patients with mild-modest renal failure and require no dose adjustments; however, in patients with severe renal impairment (creatinine clearances <30 mL/minute), data are limited on the use of sofosbuvir-containing combinations. For such patients, including those undergoing hemodialysis, recommended combinations are 12 weeks of elbasvir/grazoprevir for genotypes 1a, 1b, and 4 or 12 weeks of paritaprevir/ritonavir, ombitasvir, and dasabuvir for genotype 1b. In genotype 1a, the addition of 200 mg/day of [ribavirin](#) to paritaprevir/ritonavir, ombitasvir, and dasabuvir, if the hemoglobin level exceeds 10 g/dL, is an alternative regimen but requires vigilance for the onset of ribavirin-induced hemolytic anemia. For patients with severe renal impairment and HCV genotypes 2, 3, 5, or 6, PEG IFN with low-dose

ribavirin (200 mg daily, if the hemoglobin exceeds 10 g/dL) is recommended. After renal transplantation, levels of SVR₁₂ in patients treated with the approved oral DAA combinations have approached 100%.

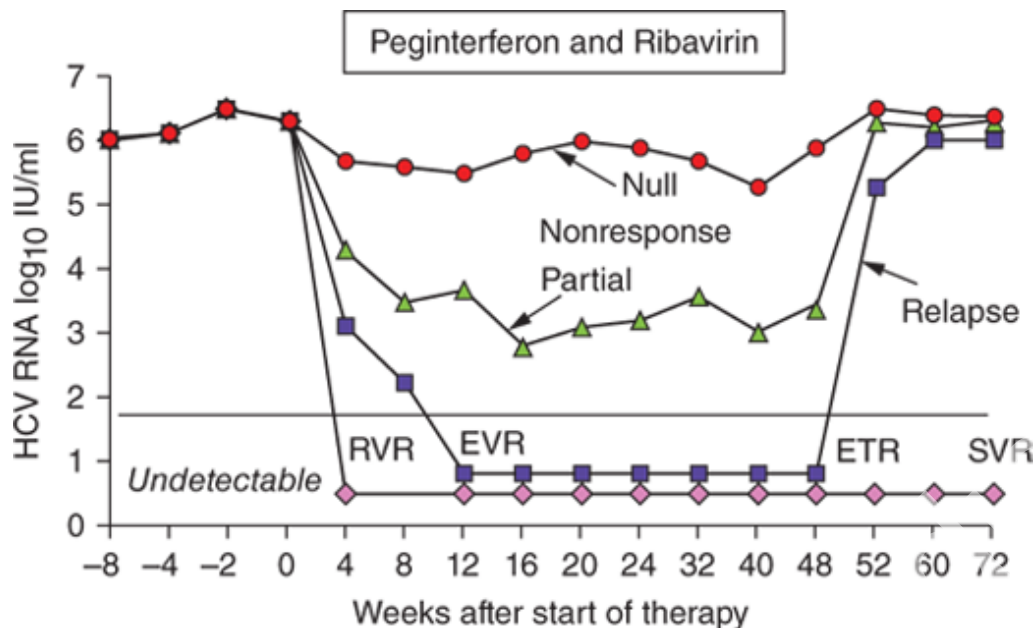
No clinical studies of the use of DAAs during pregnancy are available. **Ribavirin** is contraindicated during pregnancy; therefore, any regimen including **ribavirin** should not be used. **Sofosbuvir**; **sofosbuvir** + ledipasvir; and paritaprevir/ritonavir, ombitasvir, and dasabuvir are classified as pregnancy category B; the other DAAs do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment is compelling and justified compared to the potential for fetal risk.

Choosing among available treatment options: The large number of recommended all-oral DAA combinations can be daunting to treating clinicians. In some instances, the combination approved is determined by insurance payers; however, cost considerations aside, how is the clinician to choose among the options? The most popular of the regimens has been fixed-dose, single-pill **sofosbuvir**/ledipasvir, which is effective for all genotypes except 2 and 3, which requires no baseline RAS testing, and which can be used in noncirrhotic patients with genotype 1 and low-level viremia for as brief a period as 8 weeks. For genotypes 2 and 3, fixed-dose, single-pill **sofosbuvir**/velpatasvir appears to be the combination of choice; because this combination is so effective across all genotypes, in the future, for simplicity, clinicians may resort to a “one-size-fits-all” regimen such as this one in all patients (except for those with advanced renal failure). In addition, this regimen is the only one that can be used in almost all situations (independent of genotype, treatment experience, and cirrhosis) without **ribavirin**, and the duration of which is almost always 12 weeks; exceptions: a) **ribavirin** recommended for decompensated cirrhosis, b) EASL recommends adding **ribavirin** in treatment-experienced patients with genotype 3 or, if **ribavirin** is contraindicated, extending treatment to 24 weeks (Table 334-6, footnote c). As noted above, protease-inhibitor-containing DAA regimens (elbasvir/grazoprevir; paritaprevir/ritonavir, ombitasvir, and dasabuvir; simeprevir and **sofosbuvir**) are contraindicated in decompensated cirrhosis. For advanced renal failure, safety and efficacy have been documented for elbasvir/grazoprevir and paritaprevir/ritonavir, ombitasvir, and dasabuvir, but not for sofosbuvir-NS5A combinations.

FIGURE 334-2

Classification of virologic responses based on outcomes during and after a 48-week course of pegylated interferon (PEG IFN) plus **ribavirin antiviral therapy in patients with hepatitis C, genotype 1 or 4 (for genotype 2 or 3, the course would be 24 weeks).** Nonresponders can be classified as null responders (hepatitis C virus [HCV] RNA reduction of $<2 \log_{10}$ IU/mL) or partial responders (HCV RNA reduction $\geq 2 \log_{10}$ IU/mL but not suppressed to undetectable) by week 24 of therapy. In responders, HCV RNA can become undetectable, as shown with sensitive amplification assays, within 4 weeks (RVR, rapid virologic response); can be reduced by $\geq 2 \log_{10}$ IU/mL within 12 weeks (early virologic response, EVR; if HCV RNA is undetectable at 12 weeks, the designation is “complete” EVR); or at the end of therapy, 48 weeks (ETR, end-treatment response). In responders, if HCV RNA remains undetectable for 24 weeks after ETR, week 72, the patient has a sustained virologic response (SVR), but if HCV RNA becomes detectable again, the patient is considered to have relapsed. The posttreatment week-24 SVR (SVR₂₄) has been supplanted by an SVR at week 12 (SVR₁₂), which has been

shown to be equivalent to an SVR₂₄. In patients treated with DAA therapy, RVR and EVR milestones are largely irrelevant, being met by almost all patients. (*Reproduced with permission, courtesy of Marc G. Ghany, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health and the American Association for the Study of Liver Diseases. Hepatology 49:1335, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Indications and Recommendations for Antiviral Therapy of Chronic Hepatitis C^a**Standard Indications for Therapy**

All patients with chronic HCV infection (detectable HCV RNA, with or without elevated ALT) except for those with short life expectancies owing to comorbid conditions.

Any stage of fibrosis; highest priority for advanced fibrosis [METAVIR stage 3]/cirrhosis [METAVIR stage 4] (pretreatment biopsy is no longer embraced and has been supplanted by noninvasive measures of fibrosis, e.g., imaging to determine liver elasticity)

Responsiveness in groups previously refractory to interferon-based therapy (HIV-HCV co-infection, renal insufficiency, African American and Latino ethnicity, *IL28B* non-C haplotype, obesity, insulin resistance, hepatic decompensation, etc.) is not diminished to contemporary direct-acting oral combination regimens.

Retreatment Recommended

Relapsers, partial responders, or nonresponders after a previous course of interferon-based therapy or prior direct-acting antiviral therapy (see genotype-specific recommendations below).

Antiviral Therapy Not Recommended

Pregnancy: No clinical studies of direct-acting antivirals during pregnancy are available. **Ribavirin** is contraindicated during pregnancy; therefore, any regimen including **ribavirin** should not be used. **Sofosbuvir**; **sofosbuvir** + ledipasvir; and paritaprevir/ritonavir + ombitasvir + dasabuvir are classified as pregnancy category B, but the other direct-acting antivirals do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment outweighs the potential for fetal risk.

Therapeutic Regimens (based on AASLD-IDSA recommendations,

www.hcvguidelines.org)^b

The European Association for the Study of the Liver (EASL) issued recommendations in 2016; divergences from AASLD-IDSA recommendations are summarized as a footnote below.^c

TREATMENT-NAÏVE OR RELAPSED AFTER PRIOR PEG IFN/RIBAVIRIN** THERAPY****Genotype 1a**

ledipasvir + **sofosbuvir** 12 weeks (consider 8 weeks for noncirrhotic patients with HCV RNA $<6 \times 10^6$ IU/mL)

paritaprevir/ritonavir + **ombitasvir** + **dasabuvir** + **RBV** 12 weeks (no cirrhosis) or 24 weeks (cirrhosis)

sofosbuvir + **simeprevir** 12 weeks (no cirrhosis) or \pm **RBV** 24 weeks (cirrhosis)

daclatasvir + **sofosbuvir** 12 weeks (no cirrhosis) or \pm **RBV** 24 weeks (cirrhosis)

FAILED PRIOR PEG**IFN/**RIBAVIRIN******THERAPY, NO****CIRRHOSIS****Genotype 1a**

ledipasvir +

sofosbuvir 12 weeks

paritaprevir/ritonavir

+ **ombitasvir** +

dasabuvir + **RBV** 12

weeks

sofosbuvir +

simeprevir 12 weeks

daclatasvir +

sofosbuvir 12 weeks

grazoprevir +

elbasvir 12 weeks

(without ELB NS5A

RASs) or + **RBV** x 16

weeks (ELB NS5A

RASs)

sofosbuvir +

velpatasvir 12 weeks

Genotype 1b

ledipasvir +

sofosbuvir 12 weeks

paritaprevir/ritonavir

+ **ombitasvir** +

dasabuvir 12 weeks

sofosbuvir +

simeprevir 12 weeks

daclatasvir +

sofosbuvir 12 weeks

grazoprevir +

elbasvir 12 weeks

sofosbuvir +

velpatasvir 12 weeks

grazoprevir + elbasvir 12 weeks (no cirrhosis or cirrhosis sans ELB NS5A RASs) or + RBV × 16 weeks (ELB NS5A RASs)

sofosbuvir + velpatasvir 12 weeks

Genotype 1b

ledipasvir + **sofosbuvir** 12 weeks (consider 8 weeks for noncirrhotic patients with HCV RNA <6 × 10⁶ IU/mL)

paritaprevir/ritonavir + ombitasvir + dasabuvir 12 weeks

sofosbuvir + simeprevir 12 weeks (no cirrhosis) or ± RBV 24 weeks (cirrhosis)

daclatasvir + **sofosbuvir** 12 weeks (no cirrhosis) or ± RBV 24 weeks (cirrhosis)

grazoprevir + elbasvir 12 weeks

sofosbuvir + velpatasvir 12 weeks

Genotype 2

sofosbuvir + velpatasvir 12 weeks

daclatasvir + **sofosbuvir** (no cirrhosis) 12 weeks or 16–24 weeks (cirrhosis)

Genotype 3

sofosbuvir + velpatasvir 12 weeks

daclatasvir + **sofosbuvir** 12 weeks (no cirrhosis) or ± RBV 24 weeks (cirrhosis)

Genotype 4

sofosbuvir + velpatasvir 12 weeks

ledipasvir + **sofosbuvir** 12 weeks

paritaprevir/r + ombitasvir + RBV 12 weeks (no dasabuvir)

grazoprevir + elbasvir 12 weeks

Genotypes 5, 6

sofosbuvir + velpatasvir 12 weeks

ledipasvir + **sofosbuvir** 12 weeks

paritaprevir/ritonavir + ombitasvir + RBV 12 weeks (no dasabuvir)

grazoprevir + elbasvir 12 weeks (prior relapse) or + RBV 16 weeks (prior nonresponse)

ledipasvir + **sofosbuvir** 24 weeks

Genotypes 5, 6

sofosbuvir + velpatasvir 12 weeks

ledipasvir + **sofosbuvir** 12 weeks

Genotype 2

sofosbuvir +

velpatasvir 12 weeks

daclatasvir +

sofosbuvir 12 weeks

Genotype 3

sofosbuvir +

velpatasvir 12 weeks

daclatasvir +

sofosbuvir 12 weeks

Genotype 4

sofosbuvir +

velpatasvir 12 weeks

ledipasvir +

sofosbuvir 12 weeks

paritaprevir/r +

ombitasvir + RBV 12 weeks (no dasabuvir)

grazoprevir +

elbasvir 12 weeks

(prior relapse) or +

RBV 16 weeks (prior nonresponse)

Genotypes 5, 6

sofosbuvir +

velpatasvir 12 weeks

ledipasvir +

sofosbuvir 12 weeks

FAILED PRIOR PEG

IFN/RIBAVIRIN

THERAPY,

COMPENSATED

CIRRHOSIS

Genotype 1a

ledipasvir +

sofosbuvir + RBV 12 weeks

ledipasvir +

sofosbuvir 24 weeks

sofosbuvir +

velpatasvir 12 weeks

grazoprevir +
 elbasvir 12 weeks
 (without ELB NS5A
 RASs) or + RBV × 16
 weeks (ELB NS5A
 RASs)
 paritaprevir/ritonavir
 + ombitasvir +
 dasabuvir + RBV 24
 weeks
 sofosbuvir +
 simeprevir ± RBV 24
 weeks (no Q80K
 variant)
 daclatasvir +
 sofosbuvir ± RBV 24
 weeks
Genotype 1b
 ledipasvir +
 sofosbuvir + RBV 12
 weeks
 ledipasvir +
 sofosbuvir 24 weeks
 sofosbuvir +
 velpatasvir 12 weeks
 grazoprevir +
 elbasvir 12 weeks
 paritaprevir/ritonavir
 + ombitasvir +
 dasabuvir 12 weeks
 sofosbuvir +
 simeprevir ± RBV 24
 weeks
 daclatasvir +
 sofosbuvir ± RBV 24
 weeks
Genotype 2
 sofosbuvir +
 velpatasvir 12 weeks
 sofosbuvir +
 daclatasvir 16 or 24

weeks

Genotype 3**sofosbuvir +****velpatasvir 12 weeks**

daclatasvir +

sofosbuvir + RBV 24

weeks

Genotype 4**sofosbuvir +****velpatasvir 12 weeks**

ledipasvir +

sofosbuvir + RBV 12

weeks

FEATURES**ASSOCIATED WITH****REDUCED****RESPONSIVENESS****TO DIRECT-ACTING****ANTIVIRAL****COMBINATION****THERAPY**

Genotype and

subtype (genotype

1a less responsive

than genotype 1b for

several drugs)

Treatment

experience

Advanced fibrosis

(bridging fibrosis,
cirrhosis)

Reduced adherence

^aRapidly evolving new recommendations continue to be issued; for up-to-date treatment recommendations, please see www.hcvguidelines.org.

^bClass-I recommendations in **bold** font, all others are Class-II recommendations

^cThe following EASL recommendations differ from those of AASLD-IDSA (Please note that, although mentioned in EASL recommendations, testing for baseline RASs is not recommended routinely, but, if reliable resistance testing available, results can be used to guide therapy.):

Genotype 1

For genotype 1, simeprevir + [sofosbuvir](#) is not recommended.

For genotype 1a, treatment-experienced patients (IFN-based regimen failures) treated with [sofosbuvir](#) + ledipasvir should have weight-based [ribavirin](#) added. If reliable testing for RASs is available, [ribavirin](#) is needed only if baseline RASs are present, and, in such patients, if [ribavirin](#) is contraindicated, [sofosbuvir](#) + ledipasvir should be extended to 24 weeks.

For genotype 1b, in treatment-naïve, noncirrhotic patients receiving paritaprevir/ritonavir + ombitasvir + dasabuvir a treatment duration of 8 weeks can be considered.

For genotype 1a, in patients treatment with grazoprevir + elbasvir, EASL recommends testing for ELB RASs even in noncirrhotics. If resistance testing is not done, the level of baseline HCV RNA should determine whether [ribavirin](#) is added and the duration of therapy. If HCV RNA >800,000 IU/mL, add [ribavirin](#) and treat for 16 weeks; if HCV RNA ≤800,000 IU/mL, [ribavirin](#) is not added, and treatment for 12 weeks suffices. If baseline testing for RASs is available, patients with HCV RNA >800,000 IU/mL and detectable RASs should be treated with [ribavirin](#) for 16 weeks. Treatment without [ribavirin](#) and for 12 weeks suffices if HCV RNA ≤800,000 IU/mL even with detectable RASs or even if HCV RNA >800,000 IU/mL with undetectable RASs.

For genotype 1a, in treatment-experienced patients (IFN-based regimen failures) treated with daclatasvir + [sofosbuvir](#), follow the same recommendations described above for ledipasvir + [sofosbuvir](#) regarding the addition of [ribavirin](#).

Genotype 2

EASL recommendations are the same as those of AASLD-IDSA.

Genotype 3

For treatment-experienced patients (IFN-based regimen failures) treated with [sofosbuvir](#) + velpatasvir or [sofosbuvir](#) + daclatasvir, if testing for baseline RASs is not available, add weight-based [ribavirin](#). If resistance testing is available, [ribavirin](#) is needed only if baseline RASs are present, and, in such patients, if [ribavirin](#) is contraindicated, treatment should be extended to 24 weeks.

Genotype 4

Treatment-experienced patients (IFN-based regimen failures) treated with [sofosbuvir](#) + ledipasvir should have weight-based [ribavirin](#) added, and, in such patients, if [ribavirin](#) is contraindicated, treatment should be extended to 24 weeks.

In treatment-experienced patients (IFN-based regimen failures) treated with grazoprevir + elbasvir, if HCV RNA >800,000 IU/mL, weight-based [ribavirin](#) should be added, and treatment should be extended to 16 weeks.

EASL recommends two additional treatment options for genotype 4 (noncirrhotic or cirrhotic) that are not included in AASLD-IDSA guidelines: [sofosbuvir](#) + daclatasvir and [sofosbuvir](#) + simeprevir. For both these options, treatment-naïve patients should be treated for 12 weeks without [ribavirin](#); treatment-experienced (IFN-based regimen failures) patients should be treated with [ribavirin](#) for 12 weeks or, if [ribavirin](#) is contraindicated, without [ribavirin](#) for 24 weeks.

Genotypes 5 and 6

Treatment-experienced patients (IFN-based regimen failures) treated with [sofosbuvir](#) + ledipasvir should have weight-based [ribavirin](#) added, and, in such patients, if [ribavirin](#) is contraindicated, treatment should be extended to 24 weeks.

EASL recommends an additional treatment option for genotype 5 and 6 (noncirrhotic or cirrhotic) that is not included in AASLD-IDSA guidelines: [sofosbuvir](#) + daclatasvir. Treatment-naïve patients should be treated for 12 weeks without [ribavirin](#); treatment-experienced (IFN-based regimen failures) patients should be treated with [ribavirin](#) for 12 weeks or, if [ribavirin](#) is contraindicated, without [ribavirin](#) for 24 weeks.

Drug doses: [sofosbuvir](#) 400 mg; ledipasvir 90 mg; paritaprevir 150 mg; ritonavir 100 mg; ombitasvir 25 mg; dasabuvir 250 mg; [ribavirin](#), weight-based: 1,000 mg (<75 Kg)-1,200 mg (≥75 kg); simeprevir 150 mg; daclatasvir 60 mg; elbasvir 50 mg; grazoprevir 100 mg; velpatasvir 100 mg.

Abbreviations: AASLD, American Association for the Study of Liver Diseases; ALT, alanine aminotransferase; ELB NS5A RASs, elbasvir NS5A resistance-associated substitutions; HCV, hepatitis C virus; IFN, interferon; IDSA, Infectious Diseases Society of America; PEG IFN, pegylated interferon; IU, international units (1 IU/mL is equivalent to ~2.5 copies/mL); RASs, resistance-associated substitutions; RBV, [ribavirin](#).

AUTOIMMUNE HEPATITIS

DEFINITION

Autoimmune hepatitis is a chronic disorder characterized by continuing hepatocellular necrosis and inflammation, usually with fibrosis, which can progress to cirrhosis and liver failure. When fulfilling criteria of severity, this type of chronic hepatitis, when untreated, may have a 6-month mortality of as high as 40%. Based on contemporary estimates of the natural history of autoimmune hepatitis, the 10-year survival is 80–98% for treated and 67% for untreated patients. The prominence of extrahepatic features of autoimmunity and seroimmunologic abnormalities in this disorder supports an autoimmune process in its pathogenesis; this concept is reflected in the prior labels *lupoid* and *plasma cell hepatitis*. Autoantibodies and other typical features of autoimmunity, however, do not occur in all cases; among the broader categories of “idiopathic” or cryptogenic chronic hepatitis, many, perhaps the majority, are probably autoimmune in origin. Cases in which hepatotropic viruses, metabolic/genetic derangements (including nonalcoholic fatty liver disease), and hepatotoxic drugs have been excluded represent a spectrum of heterogeneous liver disorders of unknown cause, a proportion of which are most likely autoimmune hepatitis.

IMMUNOPATHOGENESIS

The weight of evidence suggests that the progressive liver injury in patients with autoimmune hepatitis is the result of a cell-mediated immunologic attack directed against liver cells in the setting of a loss of, or failed, immunologic tolerance for self liver antigens. In all likelihood, predisposition to autoimmunity is inherited, whereas the liver specificity of this injury is triggered by environmental (e.g., chemical, drug [e.g., minocycline], or viral) factors. For example, patients have been described in whom apparently self-limited cases of acute hepatitis A, B, or C led to autoimmune hepatitis, presumably because of genetic susceptibility or

predisposition. Evidence to support an autoimmune pathogenesis in this type of hepatitis includes the following: (1) in the liver, the histopathologic lesions are composed predominantly of cytotoxic T cells and plasma cells; (2) circulating autoantibodies (nuclear, smooth muscle, thyroid, etc.; see below), rheumatoid factor, and hyperglobulinemia are common; (3) other autoimmune disorders—such as autoimmune thyroiditis, rheumatoid arthritis, autoimmune hemolytic anemia, ulcerative colitis, membranoproliferative glomerulonephritis, juvenile diabetes mellitus, vitiligo, celiac disease, and Sjögren's syndrome—occur with increased frequency in patients and in their relatives who have autoimmune hepatitis; (4) histocompatibility haplotypes associated with autoimmune diseases, such as HLA-B1, B8, DR3, and DR4 as well as extended haplotype *DRB1*0301* and *DRB1*0401* alleles, are common in patients with autoimmune hepatitis; and (5) this type of chronic hepatitis is responsive to glucocorticoid/immunosuppressive therapy, effective in a variety of autoimmune disorders.

Cellular immune mechanisms appear to be important in the pathogenesis of autoimmune hepatitis. In vitro studies have suggested that in patients with this disorder, CD4⁺ T lymphocytes are capable of becoming sensitized to hepatocyte membrane proteins and of destroying liver cells. Molecular mimicry by cross-reacting antigens that contain epitopes similar to liver antigens is postulated to activate these T cells, which infiltrate, and result in injury to, the liver. Abnormalities of immunoregulatory control over cytotoxic lymphocytes (impaired regulatory CD4⁺CD25⁺ T cell influences) may play a role as well. Studies of genetic predisposition to autoimmune hepatitis demonstrate that certain haplotypes are associated with the disorder, as enumerated above, as are polymorphisms in cytotoxic T lymphocyte antigens (*CTLA-4*) and tumor necrosis factor α (*TNFA*2*). The precise triggering factors, genetic influences, and cytotoxic and immunoregulatory mechanisms involved in this type of liver injury remain incompletely defined.

Intriguing clues into the pathogenesis of autoimmune hepatitis come from the observation that circulating autoantibodies are prevalent in patients with this disorder. Among the autoantibodies described in these patients are antibodies to nuclei (so-called antinuclear antibodies [ANAs], primarily in a homogeneous pattern) and smooth muscle (so-called anti-smooth-muscle antibodies, directed at actin, vimentin, and skeletin), antibodies to F-actin, anti-LKM (see below), antibodies to “soluble liver antigen” (directed against a uracil-guanine-adenine transfer RNA suppressor protein), antibodies to α -actinin, and antibodies to the liver-specific asialoglycoprotein receptor (or “hepatic lectin”) and other hepatocyte membrane proteins. Although some of these provide helpful diagnostic markers, their involvement in the pathogenesis of autoimmune hepatitis has not been established.

Humoral immune mechanisms have been shown to play a role in the extrahepatic manifestations of autoimmune and idiopathic hepatitis. Arthralgias, arthritis, cutaneous vasculitis, and glomerulonephritis occurring in patients with autoimmune hepatitis appear to be mediated by the deposition of circulating immune complexes in affected tissue vessels, followed by complement activation, inflammation, and tissue injury. While specific viral antigen-antibody complexes can be identified in acute and chronic viral hepatitis, the nature of the immune complexes in autoimmune hepatitis has not been defined.

CLINICAL FEATURES

Many of the *clinical features* of autoimmune hepatitis are similar to those described for chronic viral hepatitis. The onset of disease may be insidious or abrupt; the disease may present initially like, and be confused with, acute viral hepatitis; a history of recurrent bouts of what had been labeled *acute hepatitis* is not uncommon. In approximately a quarter of patients, the diagnosis is made in the absence of symptoms, based on abnormal liver laboratory tests. A subset of patients with autoimmune hepatitis has distinct features. Such patients are predominantly young to middle-aged women with marked hyperglobulinemia and high-titer circulating ANAs. This is the group with positive lupus erythematosus (LE) preparations (initially labeled “*lupoid*” hepatitis) in whom other autoimmune features are common. Fatigue, malaise, anorexia, amenorrhea, acne, arthralgias, and jaundice are common. Occasionally, arthritis, maculopapular eruptions (including cutaneous vasculitis), erythema nodosum, colitis, pleurisy, pericarditis, anemia, azotemia, and sicca syndrome (keratoconjunctivitis, xerostomia) occur. In some patients, complications of cirrhosis, such as ascites and edema (associated with portal hypertension and hypoalbuminemia), encephalopathy, hypersplenism, coagulopathy, or variceal bleeding may bring the patient to initial medical attention.

The course of autoimmune hepatitis may be variable. In patients with mild disease or limited histologic lesions (e.g., piecemeal necrosis without bridging), progression to cirrhosis is limited, but, even in this subset, clinical monitoring is important to identify progression; up to half left untreated can progress to cirrhosis over the course of 15 years. In North America, cirrhosis at presentation is more common in African Americans than in whites. In those with severe symptomatic autoimmune hepatitis (aminotransferase levels >10 times normal, marked hyperglobulinemia, “aggressive” histologic lesions—bridging necrosis or multilobular collapse, cirrhosis), the 6-month mortality without therapy may be as high as 40%. Such severe disease accounts for only 20% of cases; the natural history of milder disease is variable, often accentuated by spontaneous remissions and exacerbations. Especially poor prognostic signs include the presence histologically of multilobular collapse at the time of initial presentation and failure of serum bilirubin to improve after 2 weeks of therapy. Death may result from hepatic failure, hepatic coma, other complications of cirrhosis (e.g., variceal hemorrhage), and intercurrent infection. In patients with established cirrhosis, HCC may be a late complication (**Chap. 78**) but occurs less frequently than in cirrhosis associated with viral hepatitis.

Laboratory features of autoimmune hepatitis are similar to those seen in chronic viral hepatitis. Liver biochemical tests are invariably abnormal but may not correlate with the clinical severity or histopathologic features in individual cases. Many patients with autoimmune hepatitis have normal serum bilirubin, alkaline phosphatase, and globulin levels with only minimal aminotransferase elevations. Serum AST and ALT levels are increased and fluctuate in the range of 100–1000 units. In severe cases, the serum bilirubin level is moderately elevated (51–171 $\mu\text{mol/L}$ [3–10 mg/dL]). Hypoalbuminemia occurs in patients with very active or advanced disease. Serum alkaline phosphatase levels may be moderately elevated or near normal. In a small proportion of patients, marked elevations of alkaline phosphatase activity occur; in such patients, clinical and laboratory features overlap with those of primary biliary cirrhosis (**Chap. 337**). The prothrombin time is often prolonged, particularly late in the disease or during active phases.

Polyclonal hypergammaglobulinemia (>2.5 g/dL) is common in autoimmune hepatitis, as is the presence of rheumatoid factor. As noted above, circulating autoantibodies are also prevalent, most characteristically ANAs in a homogeneous staining pattern. Smooth-muscle antibodies are less specific, seen just as frequently in

chronic viral hepatitis. Because of the high levels of globulins achieved in the circulation of some patients with autoimmune hepatitis, occasionally the globulins may bind nonspecifically in solid-phase binding immunoassays for viral antibodies. This has been recognized most commonly in tests for antibodies to hepatitis C virus, as noted above. In fact, studies of autoantibodies in autoimmune hepatitis have led to the recognition of new categories of autoimmune hepatitis. *Type I autoimmune hepatitis* is the classic syndrome prevalent in North America and northern Europe occurring in young women, associated with marked hyperglobulinemia, lupoid features, circulating ANAs, and HLA-DR3 or HLA-DR4 (especially *B8-DRB1*03*). Also associated with type I autoimmune hepatitis are autoantibodies against actin and atypical perinuclear antineutrophilic cytoplasmic antibodies (pANCA). Included in the spectrum of type-I autoimmune hepatitis is a subset of patients who lack ANA and anti-LKM1, but who have circulating antibodies to soluble liver antigen. Most of these patients are women and have clinical features similar to, perhaps more severe than, those of other patients with type I autoimmune hepatitis.

Type II autoimmune hepatitis, often seen in children, more common in Mediterranean populations, and linked to HLA-DRB1 and HLA-DQB1 haplotypes, is associated not with ANA but with anti-LKM. Actually, anti-LKM represent a heterogeneous group of antibodies. In type II autoimmune hepatitis, the antibody is anti-LKM1, directed against cytochrome P450 2D6. This is the same anti-LKM seen in some patients with chronic hepatitis C. Anti-LKM2 is seen in drug-induced hepatitis, and anti-LKM3 (directed against uridine diphosphate glucuronyltransferases) is seen in patients with chronic hepatitis D. Another autoantibody observed in type II autoimmune hepatitis is directed against liver cytosol formiminotransferase cyclodeaminase (anti-liver cytosol 1).

Liver biopsy abnormalities are similar to those described for chronic viral hepatitis. Expanding portal tracts and extending beyond the plate of periportal hepatocytes into the parenchyma (designated *interface hepatitis* or *piecemeal necrosis*) is a mononuclear cell infiltrate that, in autoimmune hepatitis, may include the presence of plasma cells. Necroinflammatory activity characterizes the lobular parenchyma, and evidence of hepatocellular regeneration is reflected by “rosette” formation, the occurrence of thickened liver cell plates, and regenerative “pseudolobules.” Septal fibrosis, bridging fibrosis, and cirrhosis are frequent. In patients with early autoimmune hepatitis presenting as an acute-hepatitis-like illness, lobular and centrilobular (as opposed to the more common periportal) necrosis has been reported. Bile duct injury and granulomas are uncommon; however, a subgroup of patients with autoimmune hepatitis has histologic, biochemical, and serologic features overlapping those of primary biliary cirrhosis ([Chap. 337](#)).

DIAGNOSTIC CRITERIA

An international group has suggested a set of criteria for establishing a diagnosis of autoimmune hepatitis. Exclusion of liver disease caused by genetic disorders, viral hepatitis, drug hepatotoxicity, and alcohol are linked with such inclusive diagnostic criteria as hyperglobulinemia, autoantibodies, and characteristic histologic features. This international group has also suggested a comprehensive diagnostic scoring system that, rarely required for typical cases, may be helpful when typical features are not present. Factors that weigh in favor of the diagnosis include female gender; predominant aminotransferase elevation; presence and level of globulin elevation; presence of nuclear, smooth muscle, LKM1, and other autoantibodies; concurrent other

autoimmune diseases; characteristic histologic features (interface hepatitis, plasma cells, rosettes); HLA-DR3 or DR4 markers; and response to treatment (see below). A more simplified, more specific scoring system relies on four variables: autoantibodies, serum IgG level, typical or compatible histologic features, and absence of viral hepatitis markers. Weighing against the diagnosis are predominant alkaline phosphatase elevation, mitochondrial antibodies, markers of viral hepatitis, history of hepatotoxic drugs or excessive alcohol, histologic evidence of bile duct injury, or such atypical histologic features as fatty infiltration, iron overload, and viral inclusions.

DIFFERENTIAL DIAGNOSIS

Early during the course of chronic hepatitis, autoimmune hepatitis may resemble typical *acute viral hepatitis* (**Chap. 332**). Without histologic assessment, severe chronic hepatitis cannot be readily distinguished based on clinical or biochemical criteria from mild chronic hepatitis. In adolescence, *Wilson's disease* (**Chaps. 337 and 408**) may present with features of chronic hepatitis long before neurologic manifestations become apparent and before the formation of Kayser-Fleischer rings (**copper** deposition in Descemet's membrane in the periphery of the cornea). In this age group, serum ceruloplasmin and serum and urinary **copper** determinations plus measurement of liver **copper** levels establish the correct diagnosis. *Postnecrotic or cryptogenic cirrhosis* and *primary biliary cirrhosis* (**Chap. 337**) share clinical features with autoimmune hepatitis, and both alcoholic hepatitis (**Chap. 335**) and nonalcoholic steatohepatitis (**Chap. 336**) may present with many features common to autoimmune hepatitis; historic, biochemical, serologic, and histologic assessments are usually sufficient to allow these entities to be distinguished from autoimmune hepatitis. Of course, the distinction between autoimmune and chronic viral hepatitis is not always straightforward, especially when viral antibodies occur in patients with autoimmune disease or when autoantibodies occur in patients with viral disease. Furthermore, the presence of extrahepatic features such as arthritis, cutaneous vasculitis, or pleuritis—not to mention the presence of circulating autoantibodies—may cause confusion with *rheumatologic disorders* such as rheumatoid arthritis and systemic LE. The existence of clinical and biochemical features of progressive necroinflammatory liver disease distinguishes chronic hepatitis from these other disorders, which are not associated with severe liver disease. Rarely, hepatic venous outflow obstruction (Budd-Chiari syndrome) may present with features suggestive of autoimmune hepatitis, but painful hepatomegaly, ascites, and vascular imaging provide distinguishing diagnostic clues. Other diagnostic considerations would include celiac disease and ischemic liver disease, which would be readily distinguishable by clinical and laboratory features from autoimmune hepatitis.

Finally, occasionally, features of autoimmune hepatitis overlap with features of autoimmune biliary disorders such as primary biliary cirrhosis, primary sclerosing cholangitis (**Chaps. 337 and 339**), or, even more rarely, mitochondrial antibody-negative autoimmune cholangitis. Such overlap syndromes are difficult to categorize, and often response to therapy may be the distinguishing factor that establishes the diagnosis.

TREATMENT

TREATMENT: Autoimmune Hepatitis

The mainstay of management in autoimmune hepatitis is glucocorticoid therapy. Several controlled clinical trials have documented that such therapy leads to symptomatic, clinical, biochemical, and histologic improvement as well as increased survival. A therapeutic response can be expected in up to 80% of patients. Unfortunately, therapy has not been shown in clinical trials to prevent ultimate progression to cirrhosis; however, instances of reversal of fibrosis and cirrhosis have been reported in patients responding to treatment, and rapid treatment responses within 1 year do translate into a reduction in progression to cirrhosis. Although some advocate the use of **prednisolone** (the hepatic metabolite of **prednisone**), **prednisone** is just as effective and is favored by most authorities. Therapy may be initiated at 20 mg/d, but a popular regimen in the United States relies on an initiation dose of 60 mg/d. This high dose is tapered successively over the course of a month down to a maintenance level of 20 mg/d. An alternative, but equally effective, more appealing approach is to begin with half the **prednisone** dose (30 mg/d) along with azathioprine (50 mg/d). With azathioprine maintained at 50 mg/d, the **prednisone** dose is tapered over the course of a month down to a maintenance level of 10 mg/d. The advantage of the combination approach is a reduction, over the span of an 18-month course of therapy, in serious, life-threatening complications of steroid therapy (e.g., cushingoid features, hypertension, diabetes, osteoporosis) from 66% down to under 20%. Genetic analysis for thiopurine S-methyltransferase allelic variants does not correlate with azathioprine-associated cytopenias or efficacy and is not assessed routinely in patients with autoimmune hepatitis. In combination regimens, 6-mercaptopurine may be substituted for its prodrug azathioprine, but this is rarely required. Azathioprine alone, however, is not effective in achieving remission, nor is alternate-day glucocorticoid therapy. Limited experience with **budesonide** in noncirrhotic patients suggests that this steroid side effect-sparing drug may be effective; however, the few randomized controlled trials of **budesonide** have not consistently shown efficacy. Although therapy has been shown to be effective for severe autoimmune hepatitis (AST $\geq 10 \times$ the upper limit of normal or $\geq 5 \times$ the upper limit of normal in conjunction with serum globulin greater than or equal to twice normal; bridging necrosis or multilobular necrosis on liver biopsy; presence of symptoms), therapy is not indicated for mild forms of chronic hepatitis, and the efficacy of therapy in mild or asymptomatic autoimmune hepatitis has not been established.

Improvement of fatigue, anorexia, malaise, and jaundice tends to occur within days to several weeks; biochemical improvement occurs over the course of several weeks to months, with a fall in serum bilirubin and globulin levels and an increase in serum albumin. Serum aminotransferase levels usually drop promptly, but improvements in AST and ALT alone do not appear to be reliable markers of recovery in individual patients; histologic improvement, characterized by a decrease in mononuclear infiltration and in hepatocellular necrosis, may be delayed for 6–24 months. Still, if interpreted cautiously, aminotransferase levels are valuable indicators of relative disease activity, and, although recommended, many authorities do *not* advocate for serial liver biopsies to assess therapeutic success or to guide decisions to alter or stop therapy. Rapidity of response is more common in older patients (≥ 69 years) and those with HLA *DBR1*04*; although rapid responders may progress less slowly to cirrhosis and liver transplantation, they are no less likely than slower responders to relapse after therapy. Therapy should continue for at least 12–18 months. After tapering and cessation of therapy, the likelihood of relapse is at least 50%, even if posttreatment histology has improved to show mild chronic hepatitis, and the majority of patients require therapy at maintenance doses indefinitely. Continuing azathioprine alone (2 mg/kg body weight daily) after cessation of **prednisone** therapy has been shown to

reduce the frequency of relapse. Long-term maintenance with low-dose [prednisone](#) (≤ 10 mg daily) has also been shown to keep autoimmune hepatitis in check without the theoretical risk of azathioprine marrow suppression and, in young women of child-bearing age, teratogenicity; however, maintenance azathioprine is more effective in preserving remission.

In medically refractory cases, an attempt should be made to intensify treatment with high-dose glucocorticoid monotherapy (60 mg daily) or combination glucocorticoid (30 mg daily) plus high-dose azathioprine (150 mg daily) therapy. After a month, doses of [prednisone](#) can be reduced by 10 mg a month, and doses of azathioprine can be reduced by 50 mg a month toward ultimate, conventional maintenance doses. Patients refractory to this regimen may be treated with [cyclosporine](#), [tacrolimus](#), or mycophenolate mofetil. Similarly, in exploratory studies, infusions of monoclonal antibodies directed at tumor necrosis factor (infliximab) and against the B-lymphocyte antigen CD20 (rituximab) have been reported to be of clinical benefit (improved aminotransferase levels, immunoglobulin G levels, histologic inflammatory activity) as rescue therapy for refractory autoimmune hepatitis. To date, however, only limited, often anecdotal, data in small numbers of patients support these alternative approaches. If medical therapy fails, or when chronic hepatitis progresses to cirrhosis and is associated with life-threatening complications of liver decompensation, liver transplantation is the only recourse (**Chap. 338**); in patients with severe autoimmune hepatitis, failure of the bilirubin to improve after 2 weeks of therapy should prompt early consideration of the patient for liver transplantation. Recurrence of autoimmune hepatitis in the new liver occurs rarely in most experiences but in as many as 35–40% of cases in others; nonetheless, 5-year patient and graft survival exceed 80%.

Like all patients with chronic liver disease, patients with autoimmune hepatitis should be vaccinated against hepatitis A and B, ideally before immunosuppressive therapy is begun, if practical. Patients with autoimmune hepatitis and cirrhosis should be screened for HCC with ultrasound at 6-month intervals and for gastroesophageal varices with upper gastrointestinal endoscopy at intervals of 1–3 years, based on severity of liver disease.

FURTHER READING

AASLD/IDSA HCV GUIDANCE PANEL. Hepatitis C guidance: AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus. *Hepatology* 2015;62:932–954. Updated regularly and available at <http://www.hcvguidelines.org>. Accessed September 22, 2016. [[PubMed: 26111063](#)]

Carbone M, Neuberger JM: Autoimmune liver disease, autoimmunity and liver transplantation. *J Hepatol* 60:210, 2014. [[PubMed: 24084655](#)]

European Association for the Study of the Liver: EASL clinical practice guidelines: Management of chronic hepatitis B virus infection. *J Hepatol* 57:167, 2012. [[PubMed: 22436845](#)]

European Association for the Study of the Liver: EASL clinical practice guidelines: autoimmune hepatitis. *J Hepatol* 63:971, 2015.

European Association for the Study of the Liver: EASL Recommendations on treatment of hepatitis C 2016. J Hepatol 2016 in press. Available at <http://dx.doi.org.ezproxy.uaeu.ac.ae/10.1016/j.jhep.2016.09.001>. Accessed September 22, 2016.

Lok ASG et al: Antiviral therapy for chronic hepatitis B viral infection in adults: A systematic review and meta-analysis. Hepatology 63:284, 2016. [PubMed: 26566246]

Manns MP, Czaja AJ, Gorham JD, Krawitt EL, Mieli-vergani G, Vergani D, Vierling JM. AASLD Practice Guidelines: Diagnosis and management of autoimmune hepatitis. Hepatology 51:2193, 2010.

Pawlotsky J-M et al: From non-A, non-B hepatitis to hepatitis C virus cure. J Hepatol 62:S87, 2015. [PubMed: 25920094]

Terrault N et al: AASLD guidelines for treatment of chronic hepatitis B. Hepatology 63:261, 2016. [PubMed: 26566064]

Wedemeyer H et al: Peginterferon plus adefovir versus either drug alone for hepatitis delta. N Engl J Med 364:322, 2011. [PubMed: 21268724]

ADDITIONAL ONLINE BIBLIOGRAPHY

Charlton M et al: [Ledipasvir and sofosbuvir](#) plus [ribavirin](#) for treatment of HCV infection in patients with advanced liver disease. Gastroenterology 149:649, 2015. [PubMed: 25985734]

Chhatwal J et al: Hepatitis C disease burden in the United States in the era of oral direct-acting antivirals. Hepatology 64:1442, 2016. [PubMed: 27015107]

Curry MP et al: [Sofosbuvir](#) and velpatasvir for HCV in patients with decompensated cirrhosis. N Engl J Med 373:2618, 2015. [PubMed: 26569658]

Czaja AJ: Rapidity of treatment response and outcome in type 1 autoimmune hepatitis. J Hepatol 51:161, 2009. [PubMed: 19446908]

Dienstag JL: Hepatitis B virus infection. N Engl J Med 359:1486, 2008. [PubMed: 18832247]

Feld JJ et al: [Sofosbuvir](#) and velpatasvir for HCV genotype 1, 2, 4, 5, and 6 infection. N Engl J Med 373:2599, 2015. [PubMed: 26571066]

Foster GR et al: [Sofosbuvir](#) and velpatasvir for HCV genotype 2 and 3 infection. N Engl J Med 373:2608, 2015. [PubMed: 26575258]

Foster GR et al: Impact of direct acting antiviral therapy in patients with chronic hepatitis C and decompensated cirrhosis. J Hepatol 64:1224, 2016. [[PubMed: 26829205](#)]

Heller T et al: Long-term therapy of chronic delta hepatitis with peginterferon alfa. Aliment Pharmacol Ther 40:93, 2014. [[PubMed: 24815494](#)]

Liberal R et al: Update on autoimmune hepatitis. J Clin Trans Hepatol 3:42, 2015.

Paul S et al: Hepatitis B virus reactivation and prophylaxis during solid tumor chemotherapy: A systematic review and meta-analysis. Ann Intern Med 164:30, 2016. [[PubMed: 26595058](#)]

Pawlotsky JM: Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. Gastroenterology 151:70, 2016. [[PubMed: 27080301](#)]

Perrillo RP et al: American Gastroenterological Association Institute technical review on prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. Gastroenterology 148:221, 2015. [[PubMed: 25447852](#)]

Reddy KR et al: American Gastroenterological Association Institute guideline on the prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. Gastroenterology 148:215, 2015. [[PubMed: 25447850](#)]

Rizzetto M: Hepatitis D: Thirty years after. J Hepatol 50:1043, 2009. [[PubMed: 19285743](#)]

Rosen HR: Chronic hepatitis C infection. N Engl J Med 364:2429, 2011. [[PubMed: 21696309](#)]

Sarrazin C: The importance of resistance to direct antiviral drugs in HCV infection in clinical practice. J Hepatol 64:486, 2016. [[PubMed: 26409317](#)]

Sarrazin C et al: Prevalence of resistance-associated substitutions in HCV NS5A, NS5B, or NS3 and outcomes of treatment with [ledipasvir and sofosbuvir](#). Gastroenterology 151:501, 2016. [[PubMed: 27296509](#)]

Summerskill WHJ et al: [Prednisone](#) for chronic active liver disease: Dose titration, standard dose, and combination with azathioprine compared. Gut 16:876, 1975. [[PubMed: 1104411](#)]

Van der meer AJ et al: Association between sustained virological response and all-cause mortality among patients with chronic hepatitis C and advanced hepatic fibrosis. JAMA 308:2584, 2012. [[PubMed: 23268517](#)]

Webster DP et al: Hepatitis C. Lancet 385:1124, 2015. [[PubMed: 25687730](#)]

3/14/2018

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 343: The Major Histocompatibility Complex

Gerald T. Nepom

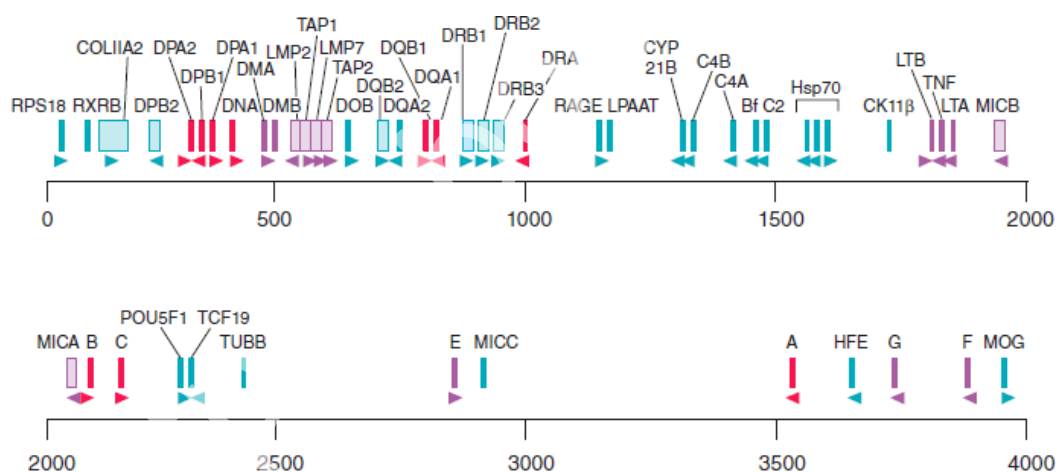
THE HLA COMPLEX AND ITS PRODUCTS

The human major histocompatibility complex (MHC), commonly called the human leukocyte antigen (HLA) complex, is a 4-megabase (Mb) region on chromosome 6 (6p21.3) that is densely packed with expressed genes. The best known of these genes are the *HLA class I* and *class II* genes, whose products are critical for immunologic specificity and transplantation histocompatibility, and they play a major role in susceptibility to a number of autoimmune diseases and some forms of drug hypersensitivity. Many other genes in the HLA region are also essential to the innate and antigen-specific functioning of the immune system. The HLA region shows extensive conservation with the MHC of other mammals in terms of genomic organization, gene sequence, and protein structure and function.

The *HLA class I* genes are located in a 2-Mb stretch of DNA at the telomeric end of the HLA region (**Fig. 343-1**). The classic (MHC class Ia) HLA-A, B, and C loci, the products of which are integral participants in the immune response to intracellular infections, tumors, and allografts, are expressed in all nucleated cells and are highly polymorphic in the population. *Polymorphism* refers to a high degree of allelic variation within a genetic locus that leads to extensive variation between different individuals expressing different alleles. More than 3400 alleles at HLA-A, 4300 alleles at HLA-B, and 3100 at HLA-C have been identified in different human populations, making this the most highly polymorphic segment known within the human genome. Each of the alleles at these loci encodes a *heavy chain* (also called an α chain) that associates noncovalently with the nonpolymorphic light chain β_2 -microglobulin, encoded on chromosome 15.

FIGURE 343-1

Physical map of the HLA region, showing the class I and class II loci, other immunologically important loci, and a sampling of other genes mapped to this region. Gene orientation is indicated by *arrowheads*. Scale is in kilobase (kb). The approximate genetic distance from DP to A is 3.2 cM. This includes 0.8 cM between A and B (including 0.2 cM between C and B), 0.4–0.8 cM between B and DR-DQ, and 1.6–2.0 cM between DR-DQ and DP.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

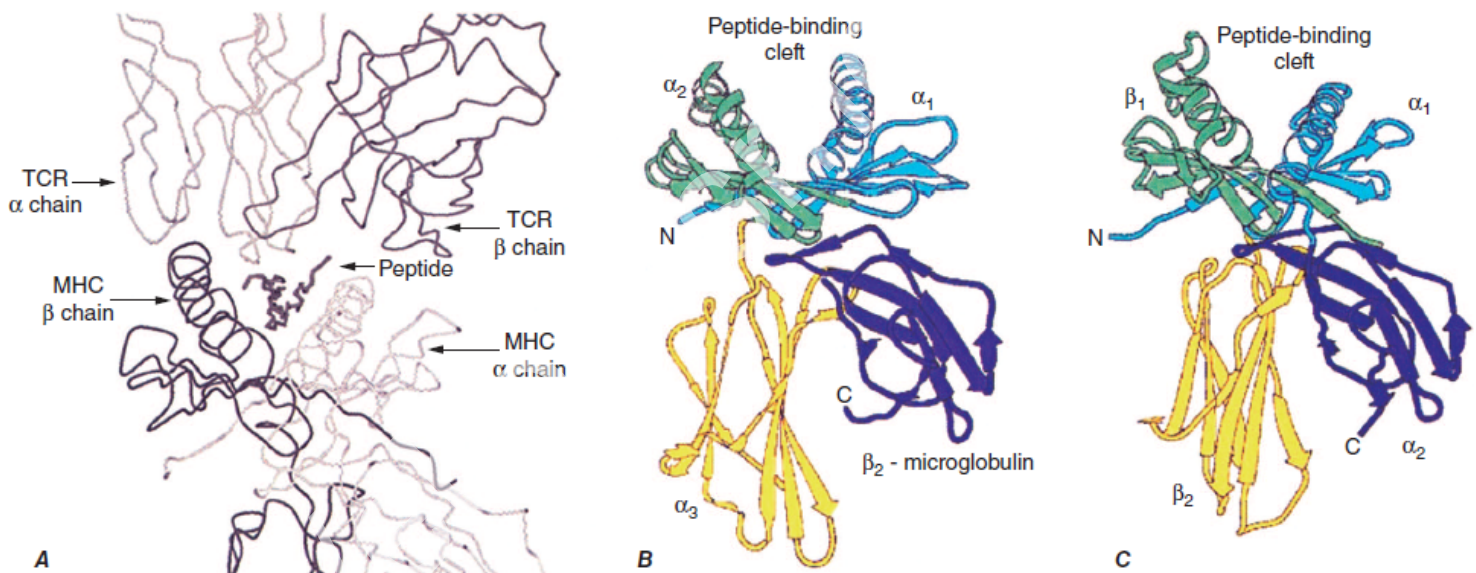
The designation of HLA genes and their products is based on a World Health Organization (WHO) nomenclature, in which alleles are given a single designation that indicates locus, allotype, and sequence-based subtype. For example, *HLA-A*02:01* indicates subtype 1 of a group of alleles that encode HLA-A2 molecules. Subtypes that differ from each other at the nucleotide but not the amino acid sequence level are designated by an extra numeral (e.g., *HLA-B*07:02:01* and *HLA-B*07:02:02* are two variants of *HLA-B*07:02*, both encoding the same HLA-B7 molecule). The nomenclature of class II genes, discussed below, is made more complicated by the fact that both chains of a class II molecule are encoded by closely linked HLA-encoded loci, each of which may be polymorphic, and by the presence of differing numbers of isotypic DRB loci in different individuals. It has become clear that accurate HLA genotyping requires DNA sequence analysis, and the identification of alleles at the

DNA sequence level has contributed greatly to the understanding of the role of HLA molecules as peptide-binding ligands, to the analysis of associations of HLA alleles with certain diseases, to the study of the population genetics of HLA, and to a clearer understanding of the contribution of HLA differences to allograft rejection and graft-versus-host disease. Current databases of HLA class I and class II sequences can be accessed by the Internet (e.g., from the IMGT/HLA Database, <http://www.ebi.ac.uk/imgt/hla>), and frequent updates of HLA gene lists are published in several journals. It is also possible to predict HLA genotypes by virtue of their linkage with single nucleotide polymorphisms (SNPs) prevalent in the genome. Imputation of HLA alleles using this technique is not as precise as targeted sequencing; however, the technology is much simpler and cheaper.

The biologic significance of this MHC genetic diversity, resulting in extreme variation in the human population, is evident from the perspective of the structure of MHC molecules. As shown in **Fig. 343-2**, the MHC class I and class II genes encode MHC molecules that bind small peptides, and together this complex (pMHC; peptide-MHC) forms the ligand for recognition by T lymphocytes, through the antigen-specific T cell receptor (TCR). There is a direct link between the genetic variation and this structural interaction: The allelic changes in genetic sequence result in diversification of the peptide-binding capabilities of each MHC molecule and in differences for specific TCR binding. Thus, different pMHC complexes bind different antigens and are targets for recognition by different T cells.

FIGURE 343-2

A. The trimolecular complex of TCR (*top*), MHC molecule (*bottom*), and a bound peptide form the structural determinants of specific antigen recognition. Other panels (**B** and **C**) show the domain structure of MHC class I (**B**) and class II (**C**) molecules. The α_1 and α_2 domains of class I and the α_1 and β_1 domains of class II form a β -sheet platform that forms the floor of the peptide-binding groove, and α helices that form the sides of the groove. The α_3 (**B**) and β_2 domains (**C**) project from the cell surface and form the contact sites for CD8 and CD4, respectively. (*Adapted from EL Reinherz et al: Science 286:1913, 1999; and C Janeway et al: Immunobiology Bookshelf, 2nd ed. Garland Publishing, New York, 1997; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The class I MHC and class II MHC structures, shown in **Fig. 343-2B,C**, are structurally closely related; however, there are a few key differences. While both bind peptides and present them to T cells, the binding pockets have different shapes, which influence the types of immune responses that result (discussed below). In addition, there are structural contact sites for T cell molecules known as CD8 and CD4, expressed on the class I or class II membrane-proximal domains, respectively. This ensures that when peptide antigens are presented by class I molecules, the responding T cells are predominantly of the CD8 class, and similarly, that T cells responding to class II pMHC complexes are predominantly CD4.

The nonclassic, or class Ib, MHC molecules, HLA-E, F, and G, are much less polymorphic than MHC Ia and appear to have distinct functions. The HLA-E molecule has a peptide repertoire displaying signal peptides cleaved from classic MHC class I molecules and is the major self-recognition target for the natural killer (NK) cell-inhibitory receptors NKG2A or NKG2C paired with CD94 (see below and **Chap. 342**). This appears to be a function of immune surveillance, because loss of MHC class I signal peptides serves as a surrogate marker for injured or infected cells, leading to release of the inhibitory signal and subsequent activation of NK cells. HLA-E can also bind and present peptides to CD8 T cells, albeit with a limited scope, as eight allelic HLA-E molecules are known. HLA-G was originally described in stem cells and in extravillous trophoblasts, where it

is implicated in regulation of maternal-fetal tolerance in pregnancy. It is now recognized as a widely expressed regulatory molecule that is expressed in multiple alternatively spliced forms, and provides inhibitory signals in both cell-bound and soluble forms; induction of expression is associated with downregulatory immunomodulation at sites of inflammation or malignancy. Eighteen allelic HLA-G molecules have been identified, interacting with receptors on NK, T cell, and dendritic cells. HLA-F occurs in four allelic forms, and is expressed on proliferating lymphoid and monocyte cells; its function is largely unknown, although it has been shown to form complexes that interact with specific NK receptors, sometimes together with other class I molecules in the absence of bound peptides. In general, the emerging view of non-classical class Ib molecules is a complex regulatory network for engaging immunomodulatory responses in the absence of traditional forms of antigen recognition attributed to classical class Ia molecules.

Additional class I-like genes have been identified, some HLA-linked and some encoded on other chromosomes, that show only distant homology to the class Ia and Ib molecules but share the three-dimensional class I structure. Those on chromosome 6p21 include MIC-A and MIC-B, which are encoded centromeric to HLA-B, and HLA-HFE, located 3 to 4 cM (centi-Morgan) telomeric of HLA-F. MIC-A and MIC-B do not bind peptide but are expressed on gut and other epithelium in a stress-inducible manner and serve as activation signals for certain $\gamma\delta$ T cells, NK cells, CD8 T cells, and activated macrophages, acting through the activating NKG2D receptors. Over 100 MIC-A and 40 MIC-B alleles are known, and additional diversification comes from variable alanine repeat sequences in the transmembrane domain. Due to this structural diversity, MIC-A can be recognized as a foreign tissue target during organ transplantation, contributing to graft failure. HLA-HFE encodes the gene defective in hereditary hemochromatosis (**Chap. 407**). Among the non-HLA, class I-like genes, CD1 refers to a family of molecules that present glycolipids or other nonpeptide ligands to certain T cells, including T cells with NK activity; FcRn binds IgG within lysosomes and protects it from catabolism (**Chap. 342**); and Zn- α_2 -glycoprotein 1 binds a nonpeptide ligand and promotes catabolism of triglycerides in adipose tissue. Like the HLA-A, B, C, E, and G heavy chains, each of which forms a heterodimer with β_2 -microglobulin (**Fig. 343-2**), the class I-like molecules, HLA-HFE, FcRn, and CD1 also bind to β_2 -microglobulin, but MIC-A, MIC-B, and Zn- α_2 -glycoprotein 1 do not.

The HLA class II region is also illustrated in **Fig. 343-1**. Multiple class II genes are arrayed within the centromeric 1 Mb of the HLA region, forming distinct haplotypes. A *haplotype* refers to an array of alleles at polymorphic loci along a chromosomal segment. Multiple class II genes are present on a single haplotype, clustered into three major subregions: HLA-DR, DQ, and DP. Each of these subregions contains at least one functional alpha (A) locus and one functional beta (B) locus. Together these encode proteins that form the α and β polypeptide chains of a mature class II HLA molecule. Thus, the DRA and DRB genes encode an HLA-DR molecule; DQA and DQB genes encode HLA-DQ molecules; and DPA and DPB genes encode HLA-DP molecules. There are several DRB genes (*DRB1*, *DRB2*, *DRB3*, etc.), so that two expressed DR molecules are encoded on most haplotypes by combining the α -chain product of the DRA gene with separate β chains. Nearly 2000 alleles have been identified at the HLA-DRB1 locus, with most of the variation occurring within limited segments encoding residues that interact with antigens. Detailed analysis of sequences and population distribution of these alleles strongly suggest that this diversity is actively selected by environmental pressures associated with pathogen diversity. In the DQ region, both DQA1 and DQB1 are polymorphic, with over 70 DQA1 alleles and 900 DQB1 alleles. The current nomenclature is largely analogous to that discussed above for class I, using the convention “locus * allele.”

In addition to allelic polymorphism, products of different DQA alleles can, with some limitations, pair with products of different DQB alleles through both *cis* and *trans* pairing to create combinatorial complexity and expand the number of expressed class II molecules. Because of the enormous allelic diversity in the general population, most individuals are heterozygous at all of the class I and class II loci. Thus, most individuals express six classic class I molecules (two each of HLA-A, -B, and -C) and many class II molecules—two DP, two to four DR, and multiple DQ (both *cis* and *trans* dimers).

OTHER GENES IN THE MHC

In addition to the class I and class II genes themselves, there are numerous genes interspersed among the HLA loci that have interesting and important immunologic functions. Our current concept of the function of MHC genes now encompasses many of these additional genes, some of which are also highly polymorphic. Indeed, direct comparison of the complete DNA sequences for eight of the entire 4-Mb MHC regions from different haplotypes shows >44,000 nucleotide variations, encoding an extremely high potential for biologic diversity, and at least 97 genes located in this region are known to have coding region sequence variation. Specific examples include the TAP and LMP genes, as discussed in more detail below, which encode molecules that participate in intermediate steps in the HLA class I biosynthetic pathway. Another set of HLA genes, DMA and DMB, performs an analogous function for the class II pathway. These genes encode an intracellular molecule that facilitates the proper complexing of HLA class II molecules with antigen (see below). The *HLA class III region* is a name given to a cluster of genes between the class I and class II complexes, which includes genes for the two closely related cytokines tumor necrosis factor (TNF)- α and lymphotoxin (TNF- β); the complement components C2, C4, and Bf; heat shock protein (HSP) 70; and the enzyme 21-hydroxylase.

The class I genes HLA-A, B, and C are expressed in all nucleated cells, although generally to a higher degree on leukocytes than on nonleukocytes. In contrast, the class II genes show a more restricted distribution: HLA-DR and HLA-DP genes are constitutively expressed on

most cells of the myeloid cell lineage, whereas all three class II gene families (HLA-DR, -DQ, and -DP) are inducible by certain stimuli provided by inflammatory cytokines such as interferon γ . Within the lymphoid lineage, expression of these class II genes is constitutive on B cells and inducible on human T cells. Most endothelial and epithelial cells in the body, including the vascular endothelium and the intestinal epithelium, are also inducible for class II gene expression, and some cells show specialized expression, such as HLA-DQA2 and HLA-DQB2 on Langerhans cells. While somatic tissues normally express only class I and not class II genes, during times of local inflammation, they are recruited by cytokine stimuli to express class II genes as well, thereby becoming active participants in ongoing immune responses. Class II expression is controlled largely at the transcriptional level through a conserved set of promoter elements that interact with a protein known as *CIITA*. Cytokine-mediated induction of *CIITA* is a principal method by which tissue-specific expression of HLA gene expression is controlled. Other HLA genes involved in the immune response, such as TAP and LMP, are also susceptible to upregulation by signals such as interferon γ .

LINKAGE DISEQUILIBRIUM

In addition to extensive polymorphism at the class I and class II loci, another characteristic feature of the HLA complex is *linkage disequilibrium*. This is formally defined as a deviation from Hardy-Weinberg equilibrium for alleles at linked loci. This is reflected in the very low recombination rates between certain loci within the HLA complex. For example, recombination between DR and DQ loci is almost never observed in family studies, and characteristic haplotypes with particular arrays of DR and DQ alleles are found in every population. Similarly, the complement components C2, C4, and Bf are almost invariably inherited together, and the alleles at these loci are found in characteristic haplotypes. In contrast, there is a recombinational hotspot between DQ and DP, which are separated by 1–2 cM of genetic distance, despite their close physical proximity. Certain extended haplotypes encompassing the interval from DQ into the class I region are commonly found, the most notable being the haplotype DR3-B8-A1, which is found, in whole or in part, in 10–30% of northern European whites. As discussed below under HLA and immunologic disease, one consequence of the phenomenon of linkage disequilibrium has been the resulting difficulty in assigning HLA-disease associations to a single allele at a single locus.

MHC STRUCTURE AND FUNCTION

Class I and class II molecules display a distinctive structural architecture, which contains specialized functional domains responsible for the unique genetic and immunologic properties of the HLA complex. The principal known function of both class I and class II HLA molecules is to bind antigenic peptides in order to present antigen to an appropriate T cell. The ability of a particular peptide to satisfactorily bind to an individual HLA molecule is a direct function of the molecular fit between the amino acid residues on the peptide with respect to the amino acid residues of the HLA molecule. The bound peptide forms a tertiary structure called the *MHC-peptide complex*, which communicates with T lymphocytes through binding to the TCR molecule. The first site of TCR-MHC-peptide interaction in the life of a T cell occurs in the thymus, where self-peptides are presented to developing thymocytes by MHC molecules expressed on thymic epithelium and hematopoietically derived antigen-presenting cells, which are primarily responsible for positive and negative selection, respectively (**Chap. 342**). Thus, the population of MHC-T cell complexes expressed in the thymus shapes the TCR repertoire. Mature T cells encounter MHC molecules in the periphery both in the maintenance of tolerance (**Chap. 348**) and in the initiation of immune responses. The TCR-MHC-peptide interaction is the central event in the initiation of most antigen-specific immune responses, since it is the structural determinant of the specificity. For potentially immunogenic peptides, the ability of a given peptide to be generated and bound by an HLA molecule is a primary feature of whether or not an immune response to that peptide can be generated, and the repertoire of peptides that a particular individual's HLA molecules can bind exerts a major influence over the specificity of that individual's immune response.

When a TCR molecule binds to an HLA-peptide complex, it forms intermolecular contacts with both the antigenic peptide and with the HLA molecule itself. The outcome of this recognition event depends on the density and duration of the binding interaction, accounting for a dual specificity requirement for activation of the T cell. That is, the TCR must be specific both for the antigenic peptide and for the HLA molecule. The polymorphic nature of the presenting molecules, and the influence that this exerts on the peptide repertoire of each molecule, results in the phenomenon of *MHC restriction* of the T cell specificity for a given peptide. The binding of CD8 or CD4 molecules to the class I or class II molecule, respectively, also contributes to the interaction between T cell and the HLA-peptide complex, by providing for the selective activation of the appropriate T cell.

CLASS I STRUCTURE

(**Fig. 343-2B**) As noted above, MHC class I molecules provide a cell-surface display of peptides derived from intracellular proteins, and they also provide the signal for self-recognition by NK cells. Surface-expressed class I molecules consist of an MHC-encoded 44-kD glycoprotein heavy chain, a non-MHC-encoded 12-kD light chain β_2 -microglobulin, and an antigenic peptide, typically 8–11 amino acids in length and derived from intracellularly produced protein. The heavy chain displays a prominent peptide-binding groove. In HLA-A and B molecules, the groove is ~3 nm in length by 1.2 nm in maximum width (30 Å \times 12 Å), whereas it is apparently somewhat wider in HLA-C. Antigenic peptides are noncovalently bound in an extended conformation within the peptide-binding groove, with both N- and C-terminal ends anchored in pockets within the

groove (A and F pockets, respectively) and, in many cases, with a prominent kink, or arch, approximately one-third of the way from the N-terminus that elevates the peptide main chain off the floor of the groove.

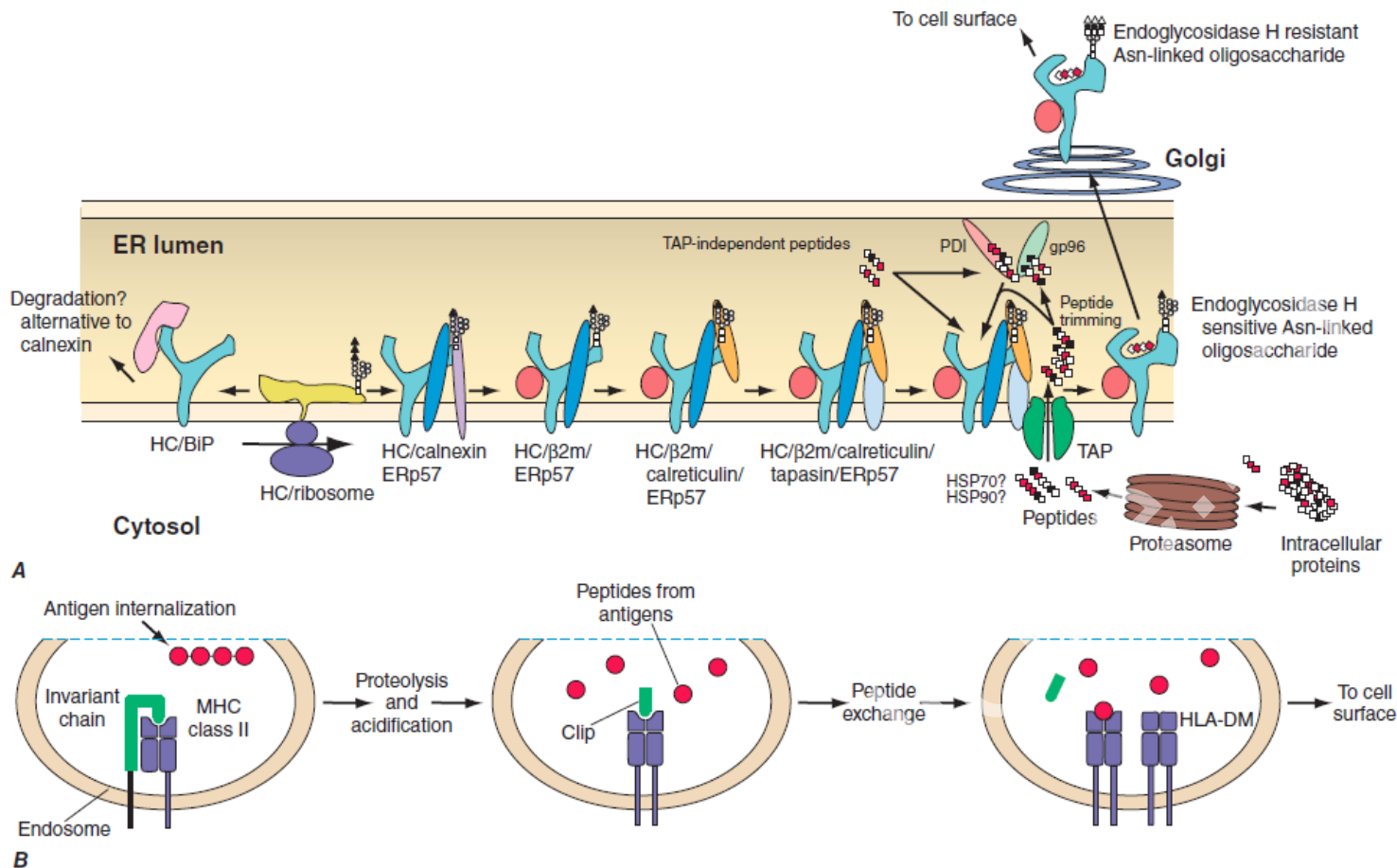
A remarkable property of peptide binding by MHC molecules is the ability to form highly stable complexes with a wide array of peptide sequences. This is accomplished by a combination of peptide sequence-independent and peptide sequence-dependent bonding. The former consists of hydrogen bond and van der Waals interactions between conserved residues in the peptide-binding groove and charged or polar atoms along the peptide backbone. The latter is dependent upon the six side pockets that are formed by the irregular surface produced by protrusion of amino acid side chains from within the binding groove. The side chains lining the pockets interact with some of the peptide side chains. The sequence polymorphism among different class I alleles and isotypes predominantly affects the residues that line these pockets, and the interactions of these residues with peptide residues constitute the sequence-dependent bonding that confers a particular sequence “motif” on the range of peptides that can bind each MHC molecule.

CLASS I BIOSYNTHESIS

(Fig. 343-3A) The biosynthesis of the classic MHC class I molecules reflects their role in presenting endogenous peptides. The heavy chain is cotranslationally inserted into the membrane of the endoplasmic reticulum (ER), where it becomes glycosylated and associates sequentially with the chaperone proteins calnexin and ERp57. It then forms a complex with β_2 -microglobulin, and this complex associates with the chaperone calreticulin and the MHC-encoded molecule tapasin, which physically links the class I complex to TAP, the MHC-encoded transporter associated with antigen processing. Meanwhile, peptides generated within the cytosol from intracellular proteins by the multisubunit, multicatalytic proteasome complex are actively transported into the ER by TAP, where they are trimmed by enzymes known as ER aminopeptidases. At this point, peptides with appropriate sequence complementarity bind specific class I molecules to form complete, folded heavy chain- β_2 -microglobulin-peptide trimer complexes. These are transported rapidly from the ER, through the *cis*- and *trans*-Golgi where the N-linked oligosaccharide is further processed, and thence to the cell surface.

FIGURE 343-3

Biosynthesis of class I (A) and class II (B) molecules. **A.** Nascent heavy chain (HC) becomes associated with β_2 -microglobulin (β_2m) and peptide through interactions with a series of chaperones. Peptides generated by the proteasome are transported into the endoplasmic reticulum (ER) by TAP. Peptides undergo N-terminal trimming in the ER and become associated with chaperones, including gp96 and PDI. Once peptide binds to HC- β_2m , the HC- β_2m -peptide trimeric complex exits the ER and is transported by the secretory pathway to the cell surface. In the Golgi, the N-linked oligosaccharide undergoes maturation, with addition of sialic acid residues. Molecules are not necessarily drawn to scale. **B.** Pathway of HLA class II molecule assembly and antigen processing. After transport through the Golgi and post-Golgi compartment, the class II-invariant chain complex moves to an acidic endosome, where the invariant chain is proteolytically cleaved into fragments and displaced by antigenic peptides, facilitated by interactions with the DMA-DMB chaperone protein. This class II molecule-peptide complex is then transported to the cell surface.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Most of the peptides transported by TAP are produced in the cytosol by proteolytic cleavage of intracellular proteins by the multisubunit, multicatalytic proteasome, and inhibitors of the proteasome dramatically reduce expression of class I-presented antigenic peptides. A thiol-dependent oxidoreductase ERp57, which mediates disulfide bond rearrangements, also appears to play an important role in folding the class I-peptide complex into a stable multicomponent molecule. The MHC-encoded proteasome subunits LMP2 and LMP7 may influence the spectrum of peptides produced but are not essential for proteasome function.

CLASS I FUNCTION

Peptide Antigen Presentation

On any given cell, a class I molecule occurs in 100,000–200,000 copies and binds several hundred to several thousand distinct peptide species. The vast majority of these peptides are self-peptides to which the host immune system is tolerant by one or more of the mechanisms that maintain tolerance (e.g., clonal deletion in the thymus or clonal anergy or clonal ignorance in the periphery [Chaps. 342 and 348]). However, class I molecules bearing foreign peptides expressed in a permissive immunologic context activate CD8 T cells, which, if naïve, will then differentiate into cytolytic T lymphocytes (CTLs). These T cells and their progeny, through their αβ TCRs, are then capable of Fas/CD95- and/or perforin-mediated cytotoxicity and/or cytokine secretion (Chap. 342) upon further encounter with the class I-peptide combination that originally activated it, or other structurally related class I-peptide complexes. As alluded to above, this phenomenon by which T cells recognize foreign antigens in the context of specific MHC alleles is termed *MHC restriction*, and the specific MHC molecule is termed the *restriction element*. The most common source of foreign peptides presented by class I molecules is viral infection, in the course of which peptides from viral proteins enter the class I pathway. The generation of a strong CTL response that destroys virally infected cells represents an important antigen-specific defense against many viral infections (Chap. 342). In the case of some viral infections—hepatitis B, for example—CTL-induced target cell apoptosis is thought to be a more important mechanism of tissue damage than any direct cytopathic effect of the virus itself. The importance of the class I pathway in the defense against viral infection is underscored by the identification of a number of viral products that interfere with the normal class I biosynthetic pathway and thus block the immunogenic expression of viral antigens.

Other examples of intracellularly generated peptides that can be presented by class I molecules in an immunogenic manner include peptides derived from nonviral intracellular infectious agents (e.g., *Listeria*, *Plasmodium*), tumor antigens, minor histocompatibility antigens, and certain autoantigens. There are also situations in which cell surface-expressed class I molecules are thought to acquire and present exogenously derived peptides.

HLA Class I Receptors and NK Cell Recognition

(**Chap. 342**) NK cells, which play an important role in innate immune responses, are activated to cytotoxicity and cytokine secretion by contact with cells that lack MHC class I expression, and NK cell activation is inhibited by cells that express MHC class I. In humans, the recognition of class I molecules by NK cells is carried out by three classes of receptor families, the killer cell-inhibitory cell receptor (KIR) family, the leukocyte Ig-like receptor (LIR) family, and the CD94/NKG2 family. The KIR family, also called CD158, is encoded on chromosome 19q13.4. KIR gene nomenclature is based on the number of domains (2D or 3D) and the presence of long (L) or short (S) cytoplasmic domains. The KIR2DL1 and S1 molecules primarily recognize alleles of HLA-C, which possess a lysine at position 80 (HLA-Cw2, -4, -5, and -6), whereas the KIR2DL2/S2 and KIR2DL3/S3 families primarily recognize alleles of HLA-C with asparagine at this position (HLA-Cw1, -3, -7, and -8). The KIR3DL1 and S1 molecules predominantly recognize HLA-B alleles that fall into the HLA-Bw4 class determined by residues 77–83 in the α_1 domain of the heavy chain, whereas the KIR3DL2 molecule is an inhibitory receptor for HLA-A*03. One of the KIR products, KIR2DL4, is known to be an activating receptor for HLA-G, and KIR3DL2 and KIR2DS4 have been described as immunoregulatory ligands interacting with HLA-F. The most common KIR haplotype in whites contains one activating KIR and six inhibitory KIR genes, although there is a great deal of diversity in the population, with >100 different combinations. It appears that most individuals have at least one inhibitory KIR for a self-HLA class I molecule, providing a structural basis for NK cell target specificity, which helps prevent NK cells from attacking normal cells. The importance of KIR-HLA interactions to many immune responses is illustrated by studies associating KIR3DL1 or S1 with multiple sclerosis (**Chap. 436**), an autoimmune disease, but also with partial protection against HIV (**Chap. 197**), in both cases consistent with a role for HLA-KIR-mediated NK activation. Studies also show an association of KIR2DS1 with protection from relapse following allogeneic bone marrow transplantation in acute myeloid leukemia when these inhibitory receptors in the donors do not recognize the recipient HLA-C.

The LIR gene family (CD85, also called ILT) is encoded centromeric of the KIR locus on 19q13.4, and it encodes a variety of inhibitory immunoglobulin-like receptors expressed on many lymphocyte and other hematopoietic lineages. Interaction of LIR-1 (ILT2) with NK or T cells inhibits activation and cytotoxicity, mediated by many different HLA class I molecules, including HLA-G. HLA-F also appears to interact with LIR molecules, although the functional context for this is not understood.

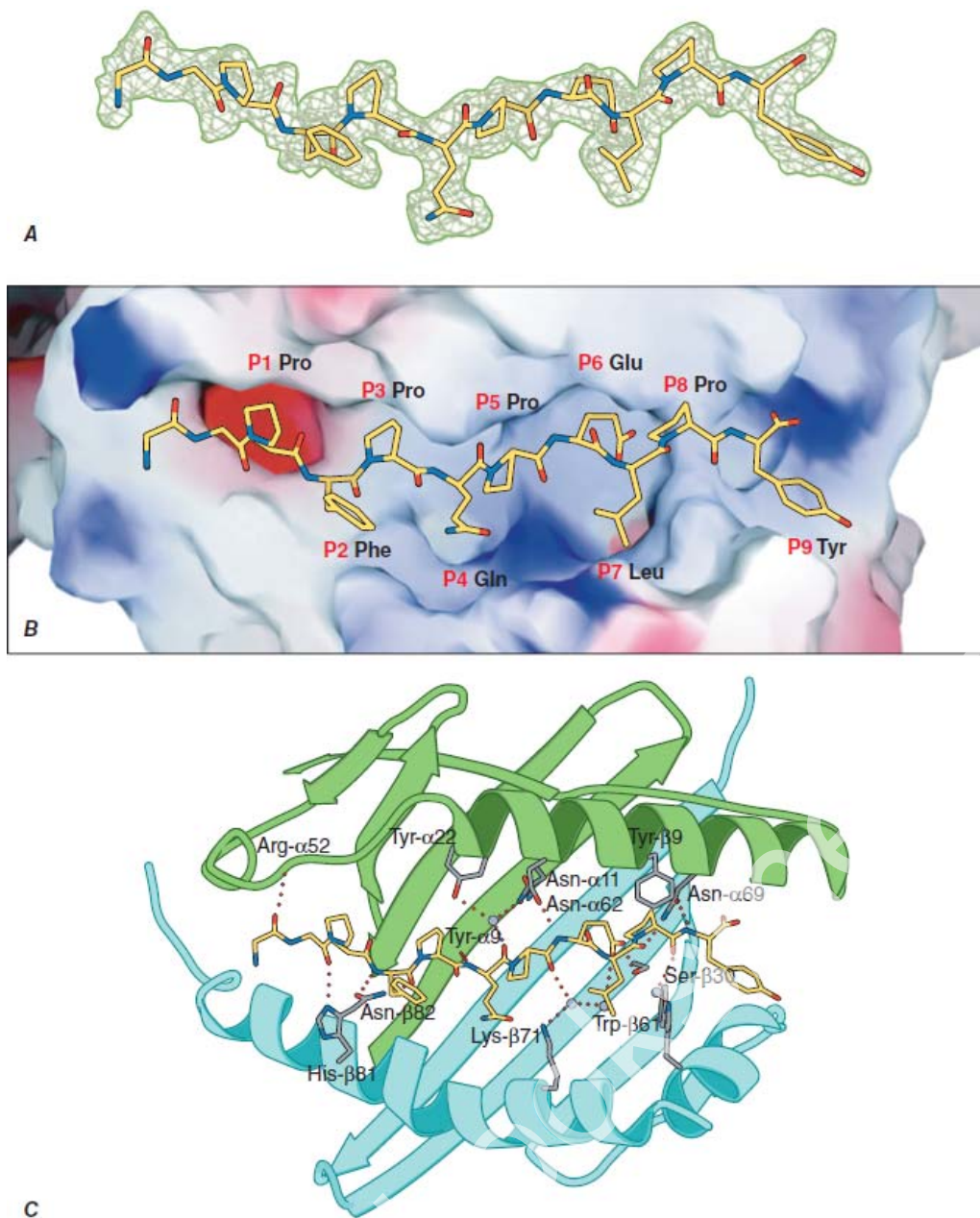
The third family of NK receptors for HLA is encoded in the NK complex on chromosome 12p12.3-13.1 and consists of CD94 and five NKG2 genes, A/B, C, E/H, D, and F. These molecules are C-type (calcium-binding) lectins, and most of these function as disulfide-bonded heterodimers between CD94 and one of the NKG2 glycoproteins. The principal ligand of CD94/NKG2A receptors is the HLA-E molecule, complexed to a peptide derived from the signal sequence of classic HLA class I molecules and HLA-G. Thus, analogous to the way in which KIR receptors recognize HLA-C, the NKG2 receptor monitors self-class I expression, albeit indirectly through peptide recognition in the context of HLA-E. NKG2C, E, and H appear to have similar specificities but act as activating receptors. NKG2D is expressed as a homodimer and functions as an activating receptor expressed on NK cells, $\gamma\delta$ TCR T cells, and activated CD8 T cells. When complexed with an adaptor called DAP10, NKG2D recognizes MIC-A and MIC-B molecules and activates the cytolytic response. NKG2D also binds a class of molecules known as *ULBP*, structurally related to class I molecules but not encoded in the MHC. **The function of NK cells in immune responses is discussed in Chap. 342.**

CLASS II STRUCTURE

(**Fig. 343-2C**) A specialized functional architecture similar to that of the class I molecules can be seen in the example of a class II molecule depicted in **Fig. 343-2C**, with an antigen-binding cleft arrayed above a supporting scaffold that extends the cleft toward the external cellular environment. However, in contrast to the HLA class I molecular structure, β_2 -microglobulin is not associated with class II molecules. Rather, the class II molecule is a heterodimer, composed of a 29-kD α chain and a 34-kD β chain. The amino-terminal domains of each chain form the antigen-binding elements that, like the class I molecule, cradle a peptide in a groove bounded by extended α -helical loops, one encoded by the A (α chain) gene and one by the B (β chain) gene. Like the class I groove, the class II antigen-binding groove is punctuated by pockets that contact the side chains of amino acid residues of the bound peptide, but unlike the class I groove, it is open at both ends. Therefore, peptides bound by class II molecules vary greatly in length, since both the N- and C-terminal ends of the peptides can extend through the open ends of this groove. Approximately 11 amino acids within the bound peptide form intimate contacts with the class II molecule itself, with backbone hydrogen bonds and specific side chain interactions combining to provide, respectively, stability and specificity to the binding (**Fig. 343-4**).

FIGURE 343-4

Specific intermolecular interactions determine peptide binding to MHC class II molecules. A short peptide sequence derived from alpha-gliadin (**A**) is accommodated within the MHC class II binding groove by specific interactions between peptide side chains (the P1–P9 residues illustrated in **B**) and corresponding pockets in the MHC class II structure. The latter are determined by the genetic polymorphisms of the MHC gene, in this case encoding an HLA-DQ2 molecule (**C**). This shows the extensive hydrogen bond and salt bridge network, which tightly constrains the pMHC complex and presents the complex of antigen and restriction element for CD4 T cell recognition. (From C Kim et al: *Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proc Natl Acad Sci USA 101:4175, 2004.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The genetic polymorphisms that distinguish different class II genes correspond to changes in the amino acid composition of the class II molecule, and these variable sites are clustered predominantly around the pocket structures within the antigen-binding groove. As with class I, this is a critically important feature of the class II molecule, which explains how genetically different individuals have functionally different HLA molecules.

BIOSYNTHESIS AND FUNCTION OF CLASS II MOLECULES

(Fig. 343-3B) The intracellular assembly of class II molecules occurs within a specialized compartmentalized pathway that differs dramatically from the class I pathway described above. As illustrated in Fig. 343-3B, the class II molecule assembles in the ER in association with a chaperone molecule, known as the *invariant chain*. The invariant chain performs at least two roles. First, it binds to the class II molecule and blocks the peptide-binding groove, thus preventing antigenic peptides from binding. This role of the invariant chain appears to account for one of the important differences between class I and class II MHC pathways, since it can explain why class I molecules present endogenous peptides from proteins newly synthesized in the ER but class II molecules generally do not. Second, the invariant chain contains molecular localization signals that direct the class II molecule to traffic into post-Golgi compartments known as *endosomes*, which develop into specialized acidic compartments where proteases cleave the invariant chain, and antigenic peptides can now occupy the class II groove. The specificity and tissue distribution of these proteases appear to be an important way in which the immune system regulates access to the peptide-binding groove and T cells become exposed to specific self-antigens. Differences in protease expression in the thymus and in the periphery may in part determine which specific peptide sequences comprise the peripheral repertoire for T cell recognition. It is at this stage in the intracellular pathway, after

cleavage of the invariant chain, that the MHC-encoded DM molecule catalytically facilitates the exchange of peptides within the class II groove to help optimize the specificity and stability of the MHC-peptide complex.

Once this MHC-peptide complex is deposited in the outer cell membrane, it becomes the target for T cell recognition via a specific TCR expressed on lymphocytes. Because the endosome environment contains internalized proteins retrieved from the extracellular environment, the class II-peptide complex often contains bound antigens that were originally derived from extracellular proteins. In this way, the class II peptide-loading pathway provides a mechanism for immune surveillance of the extracellular space. This appears to be an important feature that permits the class II molecule to bind foreign peptides, distinct from the endogenous pathway of class I-mediated presentation.

ROLE OF HLA IN TRANSPLANTATION

The development of modern clinical transplantation in the decades since the 1950s provided a major impetus for elucidation of the HLA system, as allograft survival is highest when donor and recipient are HLA-identical. Although many molecular events participate in transplantation rejection, allogeneic differences at class I and class II loci play a major role. Class I molecules can promote T cell responses in several different ways. In the cases of allografts in which the host and donor are mismatched at one or more class I loci, host T cells can be activated by classic *direct alloreactivity*, in which the antigen receptors on the host T cells react with the foreign class I molecule expressed on the allograft. In this situation, the response of any given TCR may be dominated by the allogeneic MHC molecule, the peptide bound to it, or some combination of the two. Another type of host anti-graft T cell response involves the uptake and processing of donor MHC antigens by host antigen-presenting cells and the subsequent presentation of the resulting peptides by host MHC molecules. This mechanism is termed *indirect alloreactivity*.

In the case of class I molecules on allografts that are shared by the host and the donor, a host T cell response may still be triggered because of peptides that are presented by the class I molecules of the graft but not of the host. The most common basis for the existence of these endogenous antigen peptides, called *minor histocompatibility antigens*, is a genetic difference between donor and host at a non-MHC locus encoding the structural gene for the protein from which the peptide is derived. These loci are termed *minor histocompatibility loci*, and nonidentical individuals typically differ at many such loci. CD4 T cells react to analogous class II variation, both direct and indirect, and class II differences alone are sufficient to drive allograft rejection.

ASSOCIATION OF HLA ALLELES WITH SUSCEPTIBILITY TO DISEASE

It has long been postulated that infectious agents provide the driving force for the allelic diversification seen in the HLA system. An important corollary of this hypothesis is that resistance to specific pathogens may differ between individuals, based on HLA genotype. Observations of specific HLA genes associated with resistance to malaria or dengue fever, persistence of hepatitis B, and to disease progression in HIV infection are consistent with this model. For example, failure to clear persistent hepatitis B or C viral infection may reflect the inability of particular HLA molecules to present viral antigens effectively to T cells. Similarly, both protective and susceptible HLA allelic associations have been described for human papilloma virus-associated cervical neoplasia, implicating the MHC as an influence in mediating viral clearance in this form of cancer.

Pathogen diversity is probably also the major selective pressure favoring HLA heterozygosity. The extraordinary scope of HLA allelic diversity increases the likelihood that most new pathogens will be recognized by some HLA molecules, helping to ensure immune fitness to the host. However, another consequence of diversification is that some alleles may become capable of recognition of “innocent bystander” molecules, including drugs, environmental molecules, and tissue-derived self-antigens. In a few instances, single HLA alleles display a strong selectivity for binding of a particular agent that accounts for a genetically determined response: Hypersensitivity to abacavir, an antiretroviral therapeutic, is directly linked to binding of abacavir in the antigen-binding pockets of *HLA-B*57:01*, where it is buried underneath antigenic peptides and distorts the landscape, changing T cell recognition specificity; an adverse drug reaction to abacavir is >500 times more likely to occur in persons with *HLA-B*57:01* than in individuals without this HLA allele. Other examples include chronic beryllium toxicity, which is linked to binding of beryllium by HLA-DP molecules with a specific glutamic acid polymorphic residue on the class II beta chain, clindamycin-related cutaneous drug reactions which are more common in individuals with *HLA-B*51:01*, and *dapsone* hypersensitivity in patients with leprosy who express *HLA-B*13:01*. Even in the case of more complex diseases, particular HLA alleles are strongly associated with certain inappropriate immune-mediated disease states, particularly for some common autoimmune disorders (**Chap. 348**). By comparing allele frequencies in patients with any particular disease and in control populations, >100 such associations have been identified, some of which are listed in **Table 343-1**. The strength of genetic association is reflected in the term *relative risk*, which is a statistical odds ratio representing the risk of disease for an individual carrying a particular genetic marker compared with the risk for individuals in that population without that marker. The nomenclature shown in **Table 343-1** reflects both the HLA serotype (e.g., DR3, DR4) and the HLA genotype (e.g., *DRB1*03:01*, *DRB1*04:01*). It is very likely that the class I and class II alleles themselves are the true susceptibility alleles for most of these associations. However, because of the extremely strong linkage disequilibrium between the DR and DQ loci, in some cases it has been difficult to determine the specific locus or combination of class II loci involved. In some cases, the susceptibility gene may be one of the HLA-linked genes located near the class I or class II region, but not the HLA gene itself, and in other cases, the susceptibility gene may be a non-HLA gene such as TNF- α , which is nearby. Indeed, since linkage

disequilibrium of some haplotypes extends across large segments of the MHC region, it is quite possible that combinations of genes may account for the particular associations of HLA haplotypes with disease. For example, on some haplotypes associated with rheumatoid arthritis (RA), both HLA-DRB1 alleles and a particular polymorphism associated with the TNF locus may be contributory to disease risk. Other candidates for similar epistatic effects include the IKBL gene and the MICA locus, potentially in combination with classic HLA class II risk alleles.

TABLE 343-1

Significant HLA Class I and Class II Associations with Disease

	MARKER	GENE	STRENGTH OF ASSOCIATION
Spondyloarthropathies			
Ankylosing spondylitis	B27	<i>B*27:02, -04, -05</i>	++++
Reactive arthritis (Reiter's)	B27		++++
Acute anterior uveitis	B27		+++
Reactive arthritis (<i>Yersinia</i> , <i>Salmonella</i> , <i>Shigella</i> , <i>Chlamydia</i>)	B27		+++
Psoriatic spondylitis	B27		+++
Collagen-Vascular Diseases			
Juvenile arthritis, pauciarticular	DR8	<i>DRB1*04:01, -04, -05</i>	++
Rheumatoid arthritis	DR5		++
Sjögren's syndrome	DR4		+++
Systemic lupus erythematosus	DR3		++
White	DR3		+
Japanese	DR2		++
Autoimmune Gut and Skin			
Gluten-sensitive enteropathy (celiac disease)	DQ2	<i>DQA1*05:01</i>	+++
Chronic active hepatitis	DR3	<i>DQB1*02:01</i>	++
Dermatitis herpetiformis	DR3	<i>DRB1*04:02</i>	+++
Psoriasis vulgaris	Cw6	<i>DQB1*05:03</i>	++
Pemphigus vulgaris	DR4	<i>DQB1*03:01</i>	+++
Bullous pemphigoid variant	DQ1 DQ7		+
Autoimmune Endocrine			
Type 1 diabetes mellitus	DQ8	<i>DQB1*03:02</i>	+++
Hyperthyroidism (Graves')	DR4	<i>DRB1*04:01, -04</i>	++
Hyperthyroidism (Japanese)	DR3	<i>DQB1*06:02</i>	—
Adrenal insufficiency	DR2 B8 DR3 B35 DR3		+ ^a + + ++
Autoimmune Neurologic			
Myasthenia gravis	B8	<i>DRB1*15:01</i>	+
Multiple sclerosis	DR2 DR2	<i>DRB5*01:01</i>	+ ++
Other			

	MARKER	GENE	STRENGTH OF ASSOCIATION
Behçet's disease	B51	<i>21•OH (Cyp21B)</i>	++
Congenital adrenal hyperplasia	B47	<i>DQB1*06:02</i>	+++
Narcolepsy	DR2	<i>B*57:01</i>	++++
Goodpasture's syndrome (anti-GBM)	DR2		++
Abacavir hypersensitivity	B57		++++

a

Strong negative association, that is, genetic association with protection from diabetes.

Abbreviation: GBM, glomerular basement membrane.

As might be predicted from the known function of the class I and class II gene products, almost all of the diseases associated with specific HLA alleles have an immunologic component to their pathogenesis. The recent development of soluble HLA-peptide recombinant molecules as biological probes of T cell function, often in multivalent complexes referred to as “MHC tetramers,” represents an opportunity to use HLA genetic associations to develop biomarkers for detection of early disease progression. However, it should be stressed that even the strong HLA associations with disease (those associations with relative risk of ≥ 10) implicate normal, rather than defective, alleles. Most individuals who carry these susceptibility genes do not express the associated disease; in this way, the particular HLA gene is permissive for disease but requires other environmental (e.g., the presence of specific antigens) or genetic factors for full penetrance. In each case studied, even in diseases with very strong HLA associations, the concordance of disease in monozygotic twins is higher than in HLA-identical dizygotic twins or other sibling pairs, indicating that non-HLA genes contribute to susceptibility and can significantly modify the risk attributable to HLA.

Another group of diseases is genetically linked to HLA, not because of the immunologic function of HLA alleles but rather because they are caused by autosomal dominant or recessive abnormal alleles at loci that happen to reside in or near the HLA region. Examples of these are 21-hydroxylase deficiency ([Chap. 379](#)), hemochromatosis ([Chap. 407](#)), and spinocerebellar ataxia ([Chap. 429](#)).

CLASS I ASSOCIATIONS WITH DISEASE

Although the associations of human disease with particular HLA alleles or haplotypes predominantly involve the class II region, there are also several prominent disease associations with class I alleles. These include the association of Behçet's disease ([Chap. 357](#)) with HLA-B51, psoriasis vulgaris ([Chap. 53](#)) with HLA-Cw6, and, most notably, the spondyloarthritides ([Chap. 355](#)) with HLA-B27. More than 150 HLA-B locus alleles, designated *HLA-B*27:01–B*27:154*, encode the family of B27 class I molecules. All of the subtypes share a common B pocket in the peptide-binding groove—a deep, negatively charged pocket that shows a strong preference for binding the arginine side chain. In addition, B27 is among the most negatively charged of HLA class I heavy chains, and the overall preference is for positively charged peptides. *HLA-B*27:05* is the predominant subtype in whites and most other non-Asian populations, and this subtype is very highly associated with ankylosing spondylitis (AS) ([Chap. 355](#)), both in its idiopathic form and in association with chronic inflammatory bowel disease or psoriasis vulgaris. It is also associated with reactive arthritis (ReA) ([Chap. 355](#)), with other idiopathic forms of peripheral arthritis (undifferentiated spondyloarthropathy), and with recurrent acute anterior uveitis. B27 is found in 50–90% of individuals with these conditions, compared with a prevalence of ~7% in North American whites.

Evidence that the B27 molecule itself is involved in disease pathogenesis comes both from clinical epidemiology and on the occurrence of a spondyloarthropathy-like disease in HLA-B27 transgenic rats. The association of B27 with these diseases may derive from the specificity of a particular peptide or family of peptides bound to B27 or through another mechanism that is independent of the peptide specificity of B27. In particular, HLA-B27 has been shown to form heavy chain homodimers, utilizing the cysteine residue at position 67 of the B57 α chain, in the absence of β_2 -microglobulin. These homodimers are expressed on the surface of lymphocytes and monocytes from patients with AS, and receptors including KIR3DL1, KIR3DL2, and ILT4 (LILRB2) are capable of binding to them, promoting the activation and survival of cells expressing these receptors. Alternatively, this dimerization “misfolding” of B27 may initiate an intracellular stress signaling response, called the unfolded protein response (UPR), capable of modulating immune cell function, possibly in enthesial-resident T cells that act as sensors of damage and environmental stress.

CLASS II DISEASE ASSOCIATIONS

As can be seen in [Table 343-1](#), the majority of associations of HLA and disease are with class II alleles. Several diseases have complex HLA genetic associations.

Celiac Disease

In the case of celiac disease (**Chap. 318**), it is probable that the HLA-DQ genes are the primary basis for the disease association. HLA-DQ genes present on both the celiac-associated DR3 and DR7 haplotypes include the *DQB1**02:01 gene, and further detailed studies have documented a specific class II $\alpha\beta$ dimer encoded by the *DQA1**05:01 and *DQB1**02:01 genes, which appears to account for most of the HLA genetic contribution to celiac disease susceptibility. This specific HLA association with celiac disease may have a straightforward explanation: Peptides derived from the wheat gluten component gliaden are bound to the molecule encoded by *DQA1**05:01 and *DQB1**02:01 and presented to T cells. Gliaden-derived peptides that are implicated in this immune activation bind the DQ class II dimer best when the peptide contains a glutamine to glutamic acid substitution. It has been proposed that tissue transglutaminase, an enzyme present at increased levels in the intestinal cells of celiac patients, converts glutamine to glutamic acid in gliadin, creating peptides that are capable of being bound by the DQ2 molecule and presented to T cells.

Pemphigus Vulgaris

In the case of pemphigus vulgaris (**Chap. 55**), there are two HLA genes associated with disease, *DRB1**04:02 and *DQB1**05:03. Peptides derived from desmoglein-3, an epidermal autoantigen, bind to the *DRB1**04:02- and *DQB1**05:03-encoded HLA molecules, and this combination of specific peptide binding and disease-associated class II molecule is sufficient to stimulate desmoglein-specific T cells. A bullous pemphigoid clinical variant, not involving desmoglein recognition, has been found to be associated with *HLA-DQB1**03:01.

Juvenile Arthritis

Pauciarticular juvenile arthritis (**Chap. 351**) is an autoimmune disease associated with genes at the *DRB1* locus and also with genes at the *DPB1* locus. Patients with both *DPB1**02:01 and a *DRB1* susceptibility allele (usually *DRB1**08 or *05) have a higher relative risk than expected from the additive effect of those genes alone. In juvenile patients with rheumatoid factor–positive polyarticular disease, heterozygotes carrying both *DRB1**04:01 and *04:04 have a relative risk >100, reflecting an apparent synergy in individuals inheriting both of these susceptibility genes.

Type 1 Diabetes Mellitus

Type 1 (autoimmune) diabetes mellitus (**Chap. 396**) is associated with MHC genes on more than one haplotype. The presence of both the DR3 and DR4 haplotypes in one individual confers a twentyfold increased risk for type 1 diabetes; the strongest single association is with *DQB1**03:02, and all haplotypes that carry a *DQB1**03:02 gene are associated with type 1 diabetes, whereas related haplotypes that carry a different *DQB1* gene are not. However, the relative risk associated with inheritance of this gene can be modified, depending on other HLA genes present either on the same or a second haplotype. For example, the presence of a DR2-positive haplotype containing a *DQB1**06:02 gene is associated with decreased risk. This gene, *DQB1**06:02, is considered “protective” for type 1 diabetes. Even some *DRB1* genes that can occur on the same haplotype as *DQB1**03:02 may modulate risk, so that individuals with the DR4 haplotype that contains *DRB1**04:03 are less susceptible to type 1 diabetes than individuals with other DR4-*DQB1**03:02 haplotypes. There are some characteristic structural features of the diabetes-associated DQ molecule encoded by *DQB1**03:02, particularly the capability for binding peptides that have negatively charged amino acids near their C-termini. This may indicate a role for specific antigenic peptides or T cell interactions in the immune response to islet-associated proteins.

Although the presence of a DR3 haplotype in combination with the DR4-*DQB1**0302 haplotype is a very high-risk combination for diabetes susceptibility, the specific gene on the DR3 haplotype that is responsible for this synergy is not yet identified.

Rheumatoid Arthritis

The HLA genes associated with RA (**Chap. 351**) encode a distinctive sequence of amino acids from codons 67 to 74 of the *DRβ* molecule: RA-associated class II molecules carry the sequence LeuLeuGluGlnArgArgAlaAla or LeuLeuGluGlnLysArgAlaAla in this region, whereas non-RA-associated genes carry one or more differences in this region. These residues form a portion of the molecule that lies in the middle of the α -helical portion of the *DRB1*-encoded class II molecule, termed the *shared epitope*.

The highest risk for susceptibility to RA comes in individuals who carry both a *DRB1**04:01 and *DRB1**04:04 gene. These DR4-positive RA-associated alleles with the *shared epitope* are most frequent among patients with more severe, erosive disease. Several mechanisms have been proposed that link the shared epitope to immune reactivity in RA. This portion of the class II molecule may allow preferential binding of an arthritogenic peptide, it may favor the expansion of a type of self-reactive T lymphocyte, or it may itself form part of the pMHC ligand recognized by TCR that initiates synovial tissue recognition.

MOLECULAR MECHANISMS FOR HLA-DISEASE ASSOCIATIONS

As noted above, HLA molecules play a key role in the selection and establishment of the antigen-specific T cell repertoire and a major role in the subsequent activation of those T cells during the initiation of an immune response. Precise genetic polymorphisms characteristic of individual alleles dictate the specificity of these interactions and thereby instruct and guide antigen-specific immune events. These same genetically

determined pathways are therefore implicated in disease pathogenesis when specific HLA genes are responsible for autoimmune disease susceptibility.

The fate of developing T cells within the thymus is determined by the affinity of interaction between TCR and HLA molecules bearing self-peptides, and thus the particular HLA types of each individual control the precise specificity of the T cell repertoire (**Chap. 342**). The primary basis for HLA-associated disease susceptibility may well lie within this thymic maturation pathway. The positive selection of potentially autoreactive T cells, based on the presence of specific HLA susceptibility genes, may establish the threshold for disease risk in a particular individual.

At the time of onset of a subsequent immune response, the primary role of the HLA molecule is to bind peptide and present it to antigen-specific T cells. The HLA complex can therefore be viewed as encoding genetic determinants of precise immunologic activation events. Antigenic peptides that bind particular HLA molecules are capable of stimulating T cell immune responses; peptides that do not bind are not presented to T cells and are not immunogenic. This genetic control of the immune response is mediated by the polymorphic sites within the HLA antigen-binding groove that interact with the bound peptides. In autoimmune and immune-mediated diseases, it is likely that specific tissue antigens that are targets for pathogenic lymphocytes are complexed with the HLA molecules encoded by specific susceptibility alleles. In autoimmune diseases with an infectious etiology, it is likely that immune responses to peptides derived from the initiating pathogen are bound and presented by particular HLA molecules to activate T lymphocytes that play a triggering or contributory role in disease pathogenesis. The concept that early events in disease initiation are triggered by specific HLA-peptide complexes offers some prospects for therapeutic intervention, since it may be possible to design compounds that interfere with the formation or function of specific HLA-peptide-TCR interactions.

When considering mechanisms of HLA associations with immune response and disease, it is well to remember that immune-mediated disease is a multistep process in which initial HLA-peptide recognition helps establish a repertoire of potentially reactive T cells, whereas subsequent HLA-associated antigen presentation provides the essential peptide-binding specificity for peripheral T cell recognition leading to activation. These deterministic events can occur long before clinical evidence of autoimmunity, as exemplified by the HLA genetic associations with detection of specific autoantibodies in type 1 diabetes and in rheumatoid arthritis that are present for several years before disease diagnosis.

FURTHER READING

Karlin E, Phillips E: Genotyping for severe drug hypersensitivity. *Curr Allergy Asthma Rep* 14:418, 2014. [[PubMed: 24429903](#)]

Liu J et al: Phenome-wide association study maps new diseases to the human major histocompatibility complex region. *J Med Genet* 53:681–689, 2016. [[PubMed: 27287392](#)]

Robinson J et al: The IPD and IPD-IMGT/HLA Database: Allele variant databases. *Nucleic Acids Res* 43:D423, 2015. [[PubMed: 25414341](#)]

Zheng X et al: HIBAG—HLA genotype imputation with attribute bagging. *Pharmacogenomics J* 14:192, 2014. [[PubMed: 23712092](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is 178.250.250.21

Access Provided by: National Medical Library

Silverchair

Chapter 344: Primary Immune Deficiency Diseases

Alain Fischer

INTRODUCTION

Immunity is intrinsic to life and an important tool in the fight for survival against pathogenic microorganisms. The human immune system can be divided into two major components: the innate immune system and the adaptive immune system (**Chap. 342**). The innate immune system provides the rapid triggering of inflammatory responses based on the recognition (at the cell surface or within cells) of either molecules expressed by microorganisms or molecules that serve as “danger signals” released by cells under attack. These receptor/ligand interactions trigger signaling events that ultimately lead to inflammation. Virtually all cell lineages (not just immune cells) are involved in innate immune responses; however, myeloid cells (i.e., neutrophils and macrophages) play a major role because of their phagocytic capacity. The adaptive immune system operates by clonal recognition of antigens followed by a dramatic expansion of antigen-reactive cells and execution of an immune effector program. Most of the effector cells die off rapidly, whereas memory cells persist. Although both T and B lymphocytes recognize distinct chemical moieties and execute distinct adaptive immune responses, the latter is largely dependent on the former in generating long-lived humoral immunity. Adaptive responses utilize components of the innate immune system; for example, the antigen-presentation capabilities of dendritic cells help to determine the type of effector response. Not surprisingly, immune responses are controlled by a series of regulatory mechanisms.

Hundreds of gene products have been characterized as effectors or mediators of the immune system (**Chap. 342**). Whenever the expression or function of one of these products is genetically impaired (provided the function is nonredundant), a primary immunodeficiency (PID) occurs.

PIDs are genetic diseases with primarily Mendelian inheritance. More than 350 conditions have now been described, and deleterious mutations in ~346 genes have been identified. The overall prevalence of PIDs has been estimated in various countries at 5–10 per 100,000 individuals; however, given the difficulty in diagnosing these rare and complex diseases, this figure is probably an underestimate. PIDs can involve all possible aspects of immune responses, from innate through adaptive, cell differentiation, and effector function and regulation. For the sake of clarity, PIDs should be classified according to (1) the arm of the immune system that is defective and (2) the mechanism of the defect (when known). **Table 344-1** classifies the most prevalent PIDs according to this manner of classification; however, one should bear in mind that the classification of PIDs sometimes involves arbitrary decisions because of overlap and, in some cases, lack of data.

TABLE 344-1

Classification of Primary Immune Deficiency Diseases

Deficiencies of the Innate Immune System	
<ul style="list-style-type: none"> Phagocytic cells: <ul style="list-style-type: none"> - Impaired production: severe congenital neutropenia (SCN) - Asplenia - Impaired adhesion: leukocyte adhesion deficiency (LAD) - Impaired killing: chronic granulomatous disease (CGD) Innate immunity receptors and signal transduction: <ul style="list-style-type: none"> - Defects in Toll-like receptor signaling - Mendelian susceptibility to mycobacterial disease Complement deficiencies: <ul style="list-style-type: none"> - Classical, alternative, and lectin pathways - Lytic phase 	
Deficiencies of the Adaptive Immune System	
<ul style="list-style-type: none"> T lymphocytes: <ul style="list-style-type: none"> - Impaired development - Impaired survival, migration, function 	Severe combined immune deficiencies (SCIDs) DiGeorge's syndrome Combined immunodeficiencies Hyper-IgE syndrome (autosomal dominant) DOCK3 deficiency CD40 ligand deficiency Wiskott-Aldrich syndrome Ataxia-telangiectasia and other DNA repair deficiencies
<ul style="list-style-type: none"> B lymphocytes: <ul style="list-style-type: none"> - Impaired development - Impaired function 	XL and AR agammaglobulinemia Hyper-IgM syndrome Common variable immunodeficiency (CVID) IgA deficiency
Regulatory Defects	
<ul style="list-style-type: none"> Innate immunity Adaptive immunity 	Autoinflammatory syndromes (outside the scope of this chapter) Severe colitis Hemophagocytic lymphohistiocytosis (HLH) Autoimmune lymphoproliferation syndrome (ALPS) Autoimmunity and inflammatory diseases (IPEX, APECED)

Abbreviations: APECED, autoimmune polyendocrinopathy candidiasis ectodermal dysplasia; AR, autosomal recessive; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; XL, X-linked.

The consequences of PIDs vary widely as a function of the molecules that are defective. This concept translates into multiple levels of vulnerability to infection by pathogenic and opportunistic microorganisms, ranging from extremely broad (as in severe combined immunodeficiency [SCID]) to narrowly restricted to a single microorganism (as in Mendelian susceptibility to mycobacterial disease [MSMD]). The locations of the sites of infection and the causal microorganisms involved will thus help physicians arrive at proper diagnoses. PIDs can also lead to immunopathologic responses such as allergy (as in Wiskott-Aldrich syndrome [WAS]), lymphoproliferation, and autoimmunity. A combination of recurrent infections, inflammation, and autoimmunity can be observed in a number of PIDs, thus creating obvious therapeutic challenges. Finally, some PIDs increase the risk of cancer, notably but not exclusively lymphocytic cancers, for example, lymphoma.

DIAGNOSIS OF PRIMARY IMMUNODEFICIENCIES

The most frequent symptom prompting the diagnosis of a PID is the presence of recurrent or unusually severe infections. As mentioned above, recurrent allergic or autoimmune manifestations may also alert the physician to a possible diagnosis of PID. In such cases, a detailed account of the subject's personal and family medical history should be obtained. It is of the utmost importance to gather as much medical information as possible on relatives and up to several generations of ancestors. In addition to the obvious focus on primary symptoms, the clinical examination should evaluate the size of lymphoid organs and, when appropriate, look for the characteristic signs of a number of complex syndromes that may be associated with a PID.

The performance of laboratory tests should be guided to some extent by the clinical findings. Infections of the respiratory tract (bronchi, sinuses) mostly suggest a defective antibody response. In general, invasive bacterial infections can result from complement deficiencies, signaling defects of innate immune responses, asplenia, or defective antibody responses. Viral infections, recurrent *Candida* infections, and opportunistic infections are generally suggestive of impaired T cell immunity. Skin infections and deep-seated abscesses primarily reflect innate immune defects (such as chronic granulomatous disease); however, they may also appear in the autosomal dominant hyper-IgE syndrome. [Table 344-2](#) summarizes the laboratory tests that are most frequently used to diagnose a PID. More specific tests (notably genetic tests) are then used to make a definitive diagnosis. Genomic tools now allow us to more efficiently track genetic defects through usage of gene panel resequencing and/or whole exome sequencing.

TABLE 344-2

Tests Most Frequently Used to Diagnose a Primary Immune Deficiency (PID)

TEST	INFORMATION	PID DISEASE
• Blood cell counts and cell morphology	Neutrophil counts ^a Lymphocyte counts ^a Eosinophilia Howell-Jolly bodies	↓ Severe congenital neutropenia, ↑↑ LAD T cell ID WAS, hyper-IgE syndrome Asplenia
• Chest x-ray	Thymic shadow Costochondral junctions	SCID, DiGeorge's syndrome Adenosine deaminase deficiency
• Bone x-ray	Metaphyseal ends	Cartilage hair hypoplasia
• Immunoglobulin serum levels	IgG, IgA, IgM IgE	B cell ID Hyper-IgE syndrome, WAS, T cell ID
• Lymphocyte phenotype	T, B lymphocyte counts	T cell ID, agammaglobulinemia
• Dihydrorhodamine fluorescence (DHR) assay Nitroblue tetrazolium (NBT) assay	Reactive oxygen species production by PMNs	Chronic granulomatous disease
• CH50, AP50	Classic and alternative complement pathways	Complement deficiencies
• Ultrasonography of the abdomen	Spleen size	Asplenia

a

Normal counts vary with age. For example, the lymphocyte count is between 3000 and 9000/μL of blood below the age of 3 months and between 1500 and 2500/μL in adults.

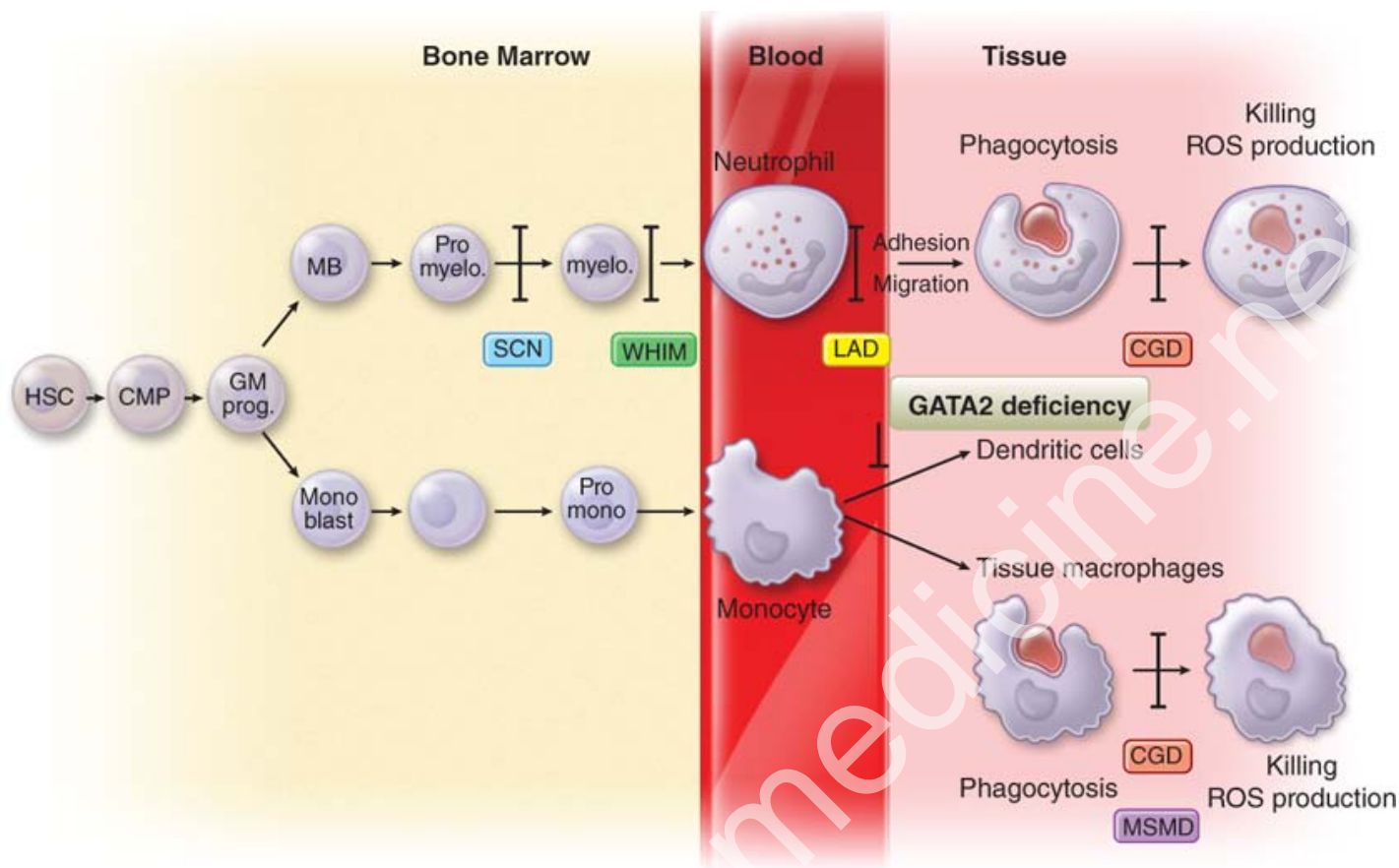
Abbreviations: ID, immunodeficiency; LAD, leukocyte adhesion deficiency; PMNs, polymorphonuclear leukocytes; SCID, severe combined immunodeficiency; WAS, Wiskott-Aldrich syndrome.

The PIDs discussed below have been grouped together according to the affected cells and the mechanisms involved ([Table 344-1](#), [Fig. 344-1](#)).

FIGURE 344-1

Differentiation of phagocytic cells and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common myeloid progenitors (CMPs) and then granulocyte-monocyte progenitors (GM-prog.), which, in turn, differentiate into neutrophils (MB: myeloblasts; Promyelo: promyelocytes; myelo: myelocytes) or monocytes (monoblasts and promonocytes). Upon activation, neutrophils adhere to the vascular endothelium, transmigrate, and phagocytose the targets. Reactive oxygen species (ROS) are delivered to the microorganism-containing phagosomes. Macrophages in tissues kill using the same mechanism. Following activation by interferon γ (not shown here), macrophages can be armed to kill intracellular pathogens such as mycobacteria. For sake of simplicity, not all cell differentiation stages are shown. The abbreviations for PIDs are contained in boxes placed at corresponding stages of the pathway. CGD, chronic granulomatous diseases; GATA2, zinc finger

transcription factor; LAD, leukocyte adhesion deficiencies; MSMD, Mendelian susceptibility to mycobacterial disease; SCN, severe congenital neutropenia; WHIM, warts, hypogammaglobulinemia, infections, and myelokathexis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

PRIMARY IMMUNODEFICIENCIES OF THE INNATE IMMUNE SYSTEM

PIDs of the innate immune system are relatively rare and account for ~10% of all PIDs.

SEVERE CONGENITAL NEUTROPENIA

Severe congenital neutropenia (SCN) consists of a group of inherited diseases that are characterized by severely impaired neutrophil counts (<500 polymorphonuclear leukocytes [PMN]/ μ L of blood). The condition is usually manifested from birth. SCN may also be cyclic (with a 3-week periodicity), and other neutropenia syndromes can also be intermittent. Although the most frequent inheritance pattern for SCN is autosomal dominant, autosomal recessive and X-linked recessive conditions also exist. Bacterial infections at the interface between the body and the external milieu (e.g., the orifices, wounds, and the respiratory tract) are common manifestations. Bacterial infections can rapidly progress through soft tissue and are followed by dissemination in the bloodstream. Severe visceral fungal infections can also ensue. The absence of pus is a hallmark of this condition.

Diagnosis of SCN requires examination of the bone marrow. Most SCNs are associated with a block in granulopoiesis at the promyelocytic stage (Fig. 344-1). SCN has multiple etiologies, and to date, mutations in 16 different genes have been identified. Most of these mutations result in isolated SCN, whereas others are syndromic (Chap. 60). The most frequent forms of SCN are caused by the premature cell death of granulocyte precursors, as observed in deficiencies of GFI1, HAX1, and elastase 2 (*ELANE*), with the latter accounting for 50% of SCN sufferers. Certain *ELANE* mutations cause cyclic neutropenia syndrome. A gain-of-function mutation in the *WASP* gene (see the section on “Wiskott-Aldrich Syndrome” below) causes X-linked SCN, which is also associated with monocytopenia.

As mentioned above, SCN exposes the patient to life-threatening, disseminated bacterial and fungal infections. Treatment requires careful hygiene measures, notably in infants. Later in life, special oral and dental care is essential, along with the prevention of bacterial infection by prophylactic administration of trimethoprim/sulfamethoxazole. Subcutaneous injection of the cytokine granulocyte colony-stimulating factor (G-CSF) usually improves neutrophil development and thus prevents infection in most SCN diseases. However, there are two caveats: (1) a few cases of SCN with *ELANE* mutation are refractory to G-CSF and may require curative treatment via allogeneic hematopoietic stem cell transplantation (HSCT); and (2) a subset of G-CSF-treated patients carrying *ELANE* mutations are at a greater risk of developing acute myelogenous leukemia associated (in most cases) with somatic gain-of-function mutations of the G-CSF receptor gene.

A few SCN conditions are associated with additional immune defects involving leukocyte migration as observed in the WHIM syndrome (gain of function mutation of the chemokine CXCR4) or in moesin deficiency.

ASPLENIA

Primary failure of the development of a spleen is an extremely rare disease that can be either syndromic (in Ivemark syndrome) or isolated with an autosomal dominant expression; in the latter case, mutations in the ribosomal protein SA and the NKX2.5 genes were recently found. Due to the absence of natural filtration of microbes in the blood, asplenia predisposes affected individuals to fulminant infections by encapsulated bacteria. Although most infections occur in the first years of life, cases may also arise in adulthood. The diagnosis is confirmed by abdominal ultrasonography and the detection of Howell-Jolly bodies in red blood cells. Effective prophylactic measures (twice-daily oral penicillin and appropriate vaccination programs) usually prevent fatal outcomes.

GATA2 DEFICIENCY

Recently an immunodeficiency combining monocytopenia and dendritic and lymphoid (B and natural killer [NK]) cell deficiency (DCML), also called monocytopenia with nontuberculous mycobacterial infections (mono-MAC), has been described as a consequence of a dominant mutation in the gene *GATA2*, a transcription factor involved in hematopoiesis. This condition also predisposes to lymphedema, myelodysplasia, and acute myeloid leukemia. Infections (bacterial and viral) are life-threatening, thus indicating, together with the malignant risk, HSCT.

LEUKOCYTE ADHESION DEFICIENCY

Leukocyte adhesion deficiency (LAD) consists of three autosomal recessive conditions (LAD I, II, and III) (**Chap. 60**). The most frequent condition (LAD I) is caused by mutations in the $\beta 2$ integrin gene; following leukocyte activation, $\beta 2$ integrins mediate adhesion to inflamed endothelium expressing cognate ligands. LAD III results from a defect in a regulatory protein (kindlin, also known as Fermt 3) involved in activating the ligand affinity of $\beta 2$ integrins. The extremely rare LAD II condition is the end result of a defect in selectin-mediated leukocyte rolling that occurs prior to $\beta 2$ integrin binding. There is a primary defect in fucose transporter such that oligosaccharide selectin ligands are missing in this syndromic condition.

Given that neutrophils are not able to reach infected tissues, LAD renders the individual susceptible to bacterial and fungal infections in a way that is similar to that of patients with SCN. LAD also causes impaired wound healing and delayed loss of the umbilical cord. A diagnosis can be suspected in cases of pus-free skin/tissue infections and massive hyperleukocytosis ($>30,000/\mu\text{L}$) in the blood (mostly granulocytes). Patients with LAD III also develop bleeding because the $\beta 2$ integrin in platelets is not functional. Use of immunofluorescence and functional assays to detect $\beta 2$ integrin can help form a diagnosis. Severe forms of LAD may require HSCT, although gene therapy is also now being considered. Neutrophil-specific granule deficiency (a very rare condition caused by a mutation in the gene for transcription factor C/EBP α) results in a condition that is clinically similar to LAD.

CHRONIC GRANULOMATOUS DISEASES

Chronic granulomatous diseases (CGDs) are characterized by impaired phagocytic killing of microorganisms by neutrophils and macrophages (**Chap. 60**). The incidence is ~ 1 per 200,000 live births. About 70% of cases are associated with X-linked recessive inheritance versus autosomal inheritance in the remaining 30%. CGD causes deep-tissue bacterial and fungal abscesses in macrophage-rich organs such as the lymph nodes, liver, and lungs. Recurrent skin infections (such as folliculitis) are common and can prompt an early diagnosis of CGD. The infectious agents are typically catalase-positive bacteria (such as *Staphylococcus*

aureus and *Serratia marcescens*) but also include *Burkholderia cepacia*, pathogenic mycobacteria (in certain regions of the world), and fungi (mainly filamentous molds, such as *Aspergillus*).

CGD is caused by defective production of reactive oxygen species (ROS) in the phagolysosome membrane following phagocytosis of microorganisms. It results from the lack of a component of NADPH oxidase (gp91phox or p22phox) or of the associated adapter/activating proteins (p47phox, p67phox, or p40phox) that mediate the transport of electrons into the phagolysosome for creating ROS by interaction with O₂. Under normal circumstances, these ROS either directly kill engulfed microorganisms or enable the rise in pH needed to activate the phagosomal proteases that contribute to microbial killing. Diagnosis of CGD is based on assays of ROS production in neutrophils and monocytes (Table 344-2). As its name suggests, CGD is also a granulomatous disease. Macrophage-rich granulomas can often arise in the liver, spleen, and other organs. These are sterile granulomas that cause disease by obstruction (bladder, pylorus, etc.) or inflammation (colitis, restrictive lung disease).

The management of infections in patients with CGD can be a complex process. The treatment of bacterial infections is generally based on combination therapy with antibiotics that are able to penetrate into cells. The treatment of fungal infections requires aggressive, long-term use of antifungals. Inflammatory/granulomatous lesions are usually steroid-sensitive; however, glucocorticoids often contribute to the spread of infections. Hence, there is strong need for new therapeutic options in what is still a poorly understood disease.

The treatment of CGD mostly relies on preventing infections. It has been unambiguously demonstrated that prophylactic usage of trimethoprim/sulfamethoxazole is both well tolerated and highly effective in reducing the risk of bacterial infection. Daily administration of azole derivatives (notably itraconazole) also reduces the frequency of fungal complications. It has long been suggested that interferon γ administration is helpful, although medical experts continue to disagree over this controversial issue. Patients may do reasonably well with prophylaxis and careful management. However, other patients develop lifelong severe and persistent fungal infections and/or chronic inflammatory complications, leading to consideration of performing HSCT. Due to increase in reported successes, HSCT is now an established curative approach for CGD; however, the risk-versus-benefit ratio must be carefully assessed on a case-by-case basis. Gene therapy approaches are also being evaluated.

MENDelian SUSCEPTIBILITY TO MYCOBACTERIAL DISEASE

This group of diseases is characterized by a defect in the interleukin-12 (IL-12)–interferon (IFN) γ axis (including IL-12p40, IL-12 receptor [R] β_1 , IFN- γ R₁ and R₂, STAT1, IRF8, and ISG515 deficiencies), which ultimately leads to impaired IFN- γ -dependent macrophage activation. Both recessive and dominant inheritance modes have been observed. The hallmark of this PID is a specific and narrow vulnerability to tuberculous and nontuberculous mycobacteria. The most severe phenotype (as observed in complete IFN- γ receptor deficiency) is characterized by disseminated infection that can be fatal even when aggressive and appropriate antimycobacterial therapy is applied. In addition to mycobacterial infections, MSMD patients (and particularly those with an IL-12/IL-12 R deficiency) are prone to developing *Salmonella* infections. Although MSMDs are very rare, they should be considered in any patient with persistent mycobacterial infection. Treatment with IFN- γ may efficiently bypass an IL-12/IL-12R deficiency.

TOLL-LIKE RECEPTOR (TLR) PATHWAY DEFICIENCIES

In a certain group of patients with early-onset, invasive *Streptococcus pneumoniae* infections or (less frequently) *Staphylococcus aureus* or other pyogenic infections, conventional screening for PIDs does not identify the cause of the defect in host defense. It has been established that these patients carry recessive mutations in genes that encode essential adaptor molecules (IRAK4 and MYD88) involved in the signaling pathways of the majority of known TLRs (Chap. 342). Remarkably, susceptibility to infection appears to decrease after the first few years of life—perhaps an indication that adaptive immunity (once triggered by an initial microbial challenge) is then able to prevent recurrent infections.

Certain TLRs (TLR-3, 7, 8, and 9) are involved in the recognition of RNA and DNA and usually become engaged during viral infections. Very specific susceptibility to herpes simplex encephalitis has been described in patients with a deficiency in Unc93b (a molecule associated with TLR-3, 7, 8, and 9 required for correct subcellular localization), TLR-3, or associated signaling molecules TRIF, TBK1, and TRAF3, resulting in defective type I IFN production. The fact that no other TLR deficiencies have been

found—despite extensive screening of patients with unexplained, recurrent infections—strongly suggests that these receptors are functionally redundant. Hypomorphic mutations in NEMO/IKK- γ (a member of the NF- κ B complex, which is activated downstream of TLR receptors) lead to a complex, variable immunodeficiency, and a number of associated features. Susceptibility to both invasive, pyogenic infections and mycobacteria may be observed in this particular setting.

COMPLEMENT DEFICIENCY

The complement system is composed of a complex cascade of plasma proteins ([Chap. 342](#)) that leads to the deposition of C3b fragments on the surface of particles and the formation of immune complexes that can culminate in the activation of a lytic complex at the bacterial surface. C3 cleavage can be mediated via three pathways: the classic, alternate, and lectin pathways. C3b coats particles as part of the opsonization process that facilitates phagocytosis following binding to cognate receptors. A deficiency in any component of the classic pathway (C1q, C1r, C1s, C4, and C2) can predispose an individual to bacterial infections that are tissue-invasive or that occur in the respiratory tract. Likewise, a C3 deficiency or a deficiency in factor I (a protein that regulates C3 consumption, thus leading to a C3 deficiency due to its absence) also results in the same type of vulnerability to infection. It has recently been reported that a very rare deficiency in ficolin-3 predisposes affected individuals to bacterial infections. Deficiencies in the alternative pathway (factors D and properdin) are associated with the occurrence of invasive *Neisseria* infections.

Lastly, deficiencies of any complement component involved in the lytic phase (C5, C6, C7, C8, and, to a lesser extent, C9) predispose affected individuals to systemic infection by *Neisseria*. This is explained by the critical role of complement in the lysis of the thick cell wall possessed by this class of bacteria.

Diagnosis of a complement deficiency relies primarily on testing the status of the classic and alternate pathway via functional assays, that is, the CH50 and AP50 tests, respectively. When either pathway is profoundly impaired, determination of the status of the relevant components in that pathway enables a precise diagnosis. Appropriate vaccinations and daily administration of oral penicillin are efficient means of preventing recurrent infections. It is noteworthy that several complement deficiencies (in the classic pathway and the lytic phase) may also predispose affected individuals to autoimmune diseases (notably systemic lupus erythematosus; [Chap. 349](#)).

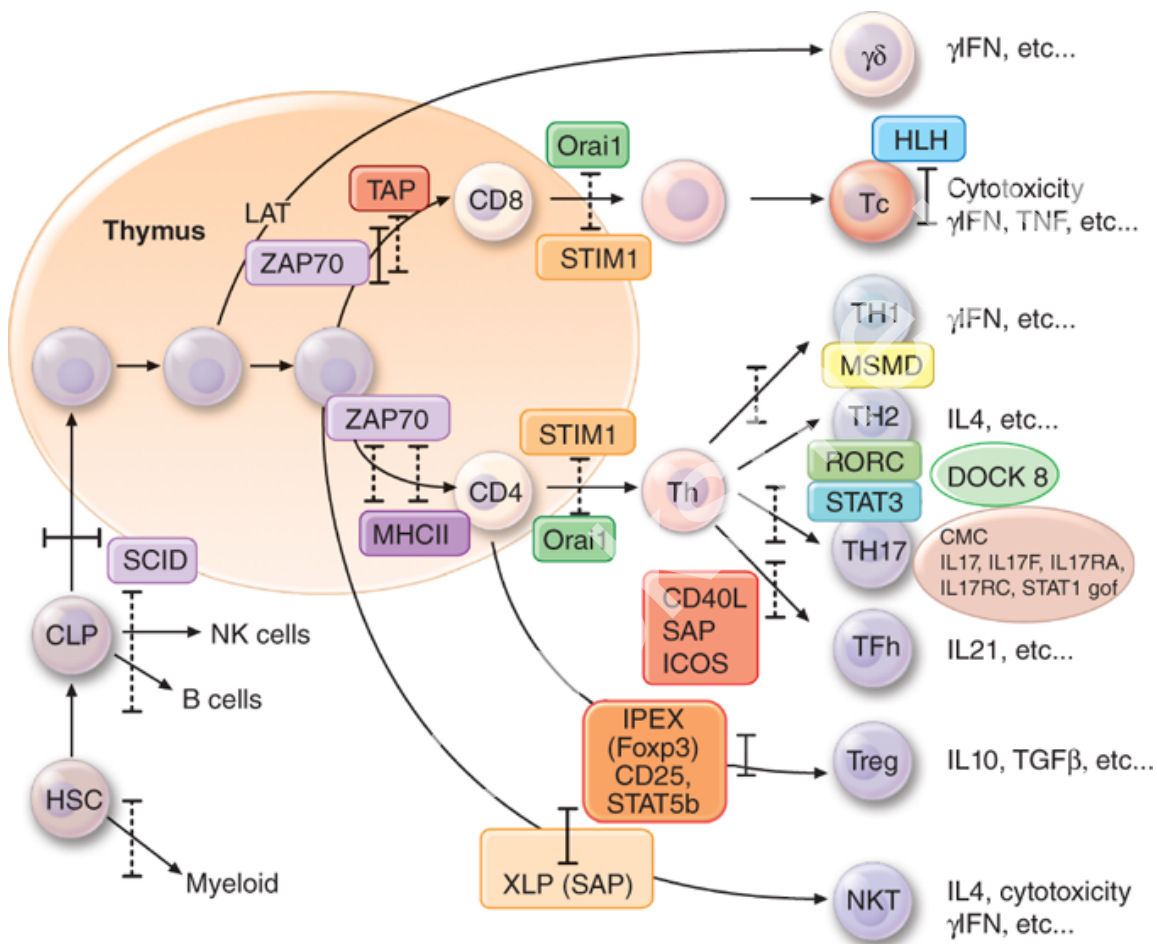
PRIMARY IMMUNODEFICIENCIES OF THE ADAPTIVE IMMUNE SYSTEM

T LYMPHOCYTE DEFICIENCIES

Given the central role of T lymphocytes in adaptive immune responses ([Chap. 342](#)), PIDs involving T cells generally have severe pathologic consequences; this explains the poor overall prognosis and the need for early diagnosis and the early intervention with appropriate therapy ([Table 344-1](#), [Figs. 344-2 and 344-3](#)). Several differentiation pathways of T cell effectors have been described, one or all of which may be affected by a given PID ([Fig. 344-2](#)). Follicular helper CD4⁺ T cells in germinal centers are required for T-dependent antibody production, including the generation of Ig class-switched, high-affinity antibodies. CD4⁺ T_H1 cells provide cytokine-dependent (mostly IFN- γ -dependent) help to macrophages for intracellular killing of various microorganisms, including mycobacteria and *Salmonella*. CD4⁺ T_H2 cells produce IL-4, IL-5, and IL-13 and thus recruit and activate eosinophils and other cells required to fight helminth infections. CD4⁺ T_H17 cells produce IL-17 and IL-22 cytokines that recruit neutrophils to the skin and lungs to fight bacterial and fungal infections. Cytotoxic CD8⁺ T cells can kill infected cells, notably in the context of viral infections. In addition, certain T cell deficiencies predispose affected individuals to *Pneumocystis jirovecii* lung infections early in life and to chronic gut/biliary duct/liver infections by *Cryptosporidium* and related genera later on in life. Lastly, naturally occurring or induced regulatory T cells are essential for controlling inflammation (notably reactivity to commensal bacteria in the gut) and autoimmunity. The role of other T cell subsets with limited T cell receptor (TCR) diversity (such as $\gamma\delta$ TCR T cells or natural killer T [NKT] cells) in PIDs is less well known; however, these subsets can be defective in certain PIDs, and this finding can sometimes contribute to the diagnosis (e.g., NKT cell deficiency in X-linked proliferative syndrome [XLP]). T cell deficiencies account for ~20% of all cases of PID.

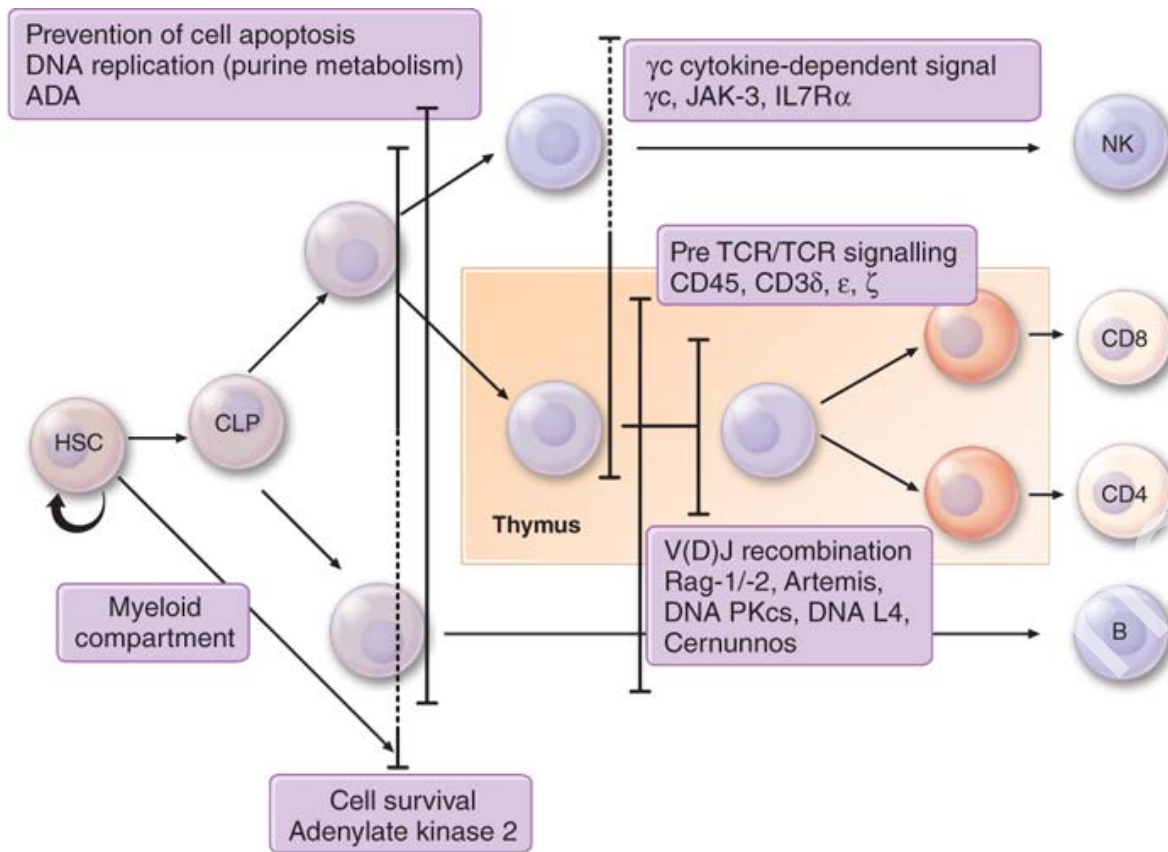
FIGURE 344-2

T cell differentiation, effector pathways, and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which, in turn, give rise to the T cell precursors that migrate to the thymus. The development of CD4⁺ and CD8⁺ T cells is shown. Known T cell effector pathways are indicated, that is, $\gamma\delta$ cells, cytotoxic T cells (Tc), T_H1, T_H2, T_H17, T_Fh (follicular helper) CD4 effector T cells, regulatory T cells (Treg), and natural killer T cells (NKTs); abbreviations for PIDs are contained in boxes. Vertical bars indicate a complete deficiency; broken bars a partial deficiency. DOCK8, autosomal recessive form of hyper-IgE syndrome; HLH, hematopoietic lymphohistiocytosis; IL17F, IL17RA, STAT1 (gof: gain of function), CMC (chronic mucocutaneous candidiasis), CD40L, ICOS, SAP deficiencies; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; LAT, linker for activation of T cells; MHCII, major histocompatibility complex class II deficiency; MSMD, Mendelian susceptibility to mycobacterial disease; Orai1, STIM1 deficiencies; RORC, RAR related orphan receptor C; SCID, severe combined immunodeficiency; STAT3, autosomal dominant form of hyper-IgE syndrome; TAP, TAP1 and TAP2 deficiencies; XLP, X-linked proliferative syndromes; ZAP70, zeta-associated protein deficiency.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE 344-3

T cell differentiation and severe combined immunodeficiencies (SCIDs). The vertical bars indicate the five mechanisms currently known to lead to SCID. The names of deficient proteins are indicated in the boxes adjacent to the vertical bars. A broken line means that deficiency is partial or involves only some of the indicated immunodeficiencies. ADA, adenosine deaminase deficiency; CLPs, common lymphoid progenitors; DNAL4, DNA ligase 4; HSCs, hematopoietic stem cells; NKs, natural killer cells; TCR, T cell receptor.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Severe Combined Immunodeficiencies

SCIDs constitute a group of rare PIDs characterized by a profound block in T cell development and thus the complete absence of these cells. The developmental block is always the consequence of an intrinsic deficiency. The incidence of SCID is estimated to be 1 in 50,000 live births. Given the severity of the T cell deficiency, clinical consequences occur early in life (usually within 3 to 6 months of birth). The most frequent clinical manifestations are recurrent oral candidiasis, failure to thrive, and protracted diarrhea and/or acute interstitial pneumonitis caused by *Pneumocystis jiroveci* (although the latter can also be observed in the first year of life in children with B cell deficiencies). Severe viral infections or invasive bacterial infections can also occur. Patients may also experience complications related to infections caused by live vaccines (notably bacille Calmette-Guérin [BCG]) that may lead not only to local and regional infection but also to disseminated infection manifested by fever, splenomegaly, and skin and lytic bone lesions. A scaly skin eruption can be observed in a context of maternal T cell engraftment (see below). A diagnosis of SCID can be suspected based on the patient's clinical history and, possibly, a family history of deaths in very young children (suggestive of either X-linked or recessive inheritance). Lymphocytopenia is strongly suggestive of SCID in >90% of cases (Table 344-2). The absence of a thymic shadow on a chest x-ray can also be suggestive of SCID. An accurate diagnosis relies on precise determination of the number of circulating T, B, and NK lymphocytes and their subsets. T cell lymphopenia may be masked in some patients by the presence of maternal T cells (derived from maternal-fetal blood transfers) that cannot be eliminated. Although counts are usually low (<500/μL of blood), higher maternal T cell counts may, under some circumstances, initially mask the presence of SCID. Thus, screening for maternal cells by using adequate genetic markers should be performed whenever necessary. Inheritance pattern analysis and lymphocyte phenotyping can discriminate between various forms of SCID and provide guidance in the choice of accurate molecular diagnostic tests (see below). To date, five distinct causative mechanisms for SCID (Fig. 344-3) have been identified. T cell quantification of receptor excision circles (TREC) by using the Guthrie card is a reliable diagnostic test for newborn screening. It is now operational in most of the United States and is being evaluated elsewhere. Its more widespread use will lead to the provision of therapy (see below) to uninfected patients resulting in a maximal chance of cure.

SEVERE COMBINED IMMUNODEFICIENCY CAUSED BY A CYTOKINE-SIGNALING DEFICIENCY

The most frequent SCID phenotype (accounting for 40–50% of all cases) is the absence of both T and NK cells. This outcome results from a deficiency in either the common γ chain (γ_c) receptor that is shared by several cytokine receptors (the IL-2, 4, 7, 9, 15, and 21 receptors) or Jak-associated kinase (JAK) 3 that binds to the cytoplasmic portion of the γ_c chain receptor and induces signal transduction following cytokine binding. The former form of SCID (γ_c deficiency) has an X-linked inheritance mode, while the second is autosomal recessive. A lack of the IL-7R α chain (which, together with γ_c , forms the IL-7 receptor) induces a selective T cell deficiency.

PURINE METABOLISM DEFICIENCY

Ten to 20% of SCID patients exhibit a deficiency in adenosine deaminase (ADA), an enzyme of purine metabolism that deaminates adenosine (ado) and deoxyadenosine (dAdo). An ADA deficiency results in the accumulation of ado and dAdo metabolites that induce premature cell death of lymphocyte progenitors. The condition results in the absence of B and NK lymphocytes as well as T cells. The clinical expression of complete ADA deficiency typically occurs very early in life. Since ADA is a ubiquitous enzyme, its deficiency can also cause bone dysplasia with abnormal costochondral junctions and metaphyses (found in 50% of cases) and neurologic defects. The very rare purine nucleoside phosphorylase (PNP) deficiency causes a profound although incomplete T cell deficiency that is often associated with severe neurologic impairments.

DEFECTIVE REARRANGEMENTS OF T AND B CELL RECEPTORS

A series of SCID conditions are characterized by a selective deficiency in T and B lymphocytes with autosomal recessive inheritance. These conditions account for 20–30% of SCID cases and result from mutations in genes encoding proteins that mediate the recombination of V(D)J gene elements in T and B cell antigen receptor genes (required for the generation of diversity in antigen recognition). The main deficiencies involve RAG1, RAG2, DNA-dependent protein kinase, and Artemis. A less severe (albeit variable) immunologic phenotype can result from other deficiencies in the same pathway, that is, DNA ligase 4 and Cernunnos deficiencies. Given that these latter factors are involved in DNA repair, these deficiencies also cause developmental defects.

DEFECTIVE (PRE-)T CELL RECEPTOR SIGNALING IN THE THYMUS

A selective T cell defect can be caused by a series of rare deficiencies in molecules involved in signaling via the pre-TCR or the TCR. These include deficiencies in CD3 subunits associated with the (pre-)TCR (i.e., CD3 δ , ϵ , and ζ) and CD45.

RETICULAR DYSGENESIS

Reticular dysgenesis is an extremely rare form of SCID that causes T and NK deficiencies with severe neutropenia and sensorineural deafness. It results from an adenylate kinase 2 deficiency.

Patients with SCID require appropriate care with aggressive anti-infective therapies, immunoglobulin replacement, and (when necessary) parenteral nutrition support. In most cases, curative treatment relies on HSCT. Today, HSCT provides a very high curative potential for SCID patients who are otherwise in reasonably good condition. In this regard, neonatal screening, based on quantification of TRECs on a Guthrie card sample, is being developed. Gene therapy has been found to be successful for cases of X-linked SCID (γ_c deficiency) and SCID caused by an ADA deficiency, although toxicity has become an issue in the treatment of the former disease that may now be overcome by use of newly generated vectors. Lastly, a third option for the treatment of ADA deficiency consists of enzyme substitution with a pegylated enzyme.

Thymic Defects

A profound T cell defect can also result from faulty development of the thymus, as is most often observed in rare cases of DiGeorge's syndrome—a relatively common condition leading to a constellation of developmental defects. In ~1% of such cases, the thymus is completely absent, leading to virtually no mature T cells. However, expansion of oligoclonal T cells can occur and is associated with skin lesions. Diagnosis (using immunofluorescence in situ hybridization) is based on the identification of a hemizygous deletion in the long arm of chromosome 22. To recover the capability for T cell differentiation, these cases require a thymic graft. CHARGE (coloboma of the eye, heart anomaly, choanal atresia, retardation, genital, and ear anomalies) syndrome (CHD7 deficiency) is a less frequent cause of impaired thymus development. Lastly, the very rare “nude” defect is characterized by the absence of both hair and the thymus.

Omenn Syndrome

Omenn syndrome consists of a subset of T cell deficiencies that present with a unique phenotype, including early-onset erythrodermia, alopecia, hepatosplenomegaly, and failure to thrive. These patients usually display T cell lymphocytosis, eosinophilia, and low B cell counts. It has been found that the T cells of these patients exhibit a low TCR heterogeneity. This peculiar syndrome is the consequence of hypomorphic mutations in genes usually associated with SCID, that is, *RAG-1*, *RAG-2*, or (less frequently) *Artemis* or *IL-7R α* . The impaired homeostasis of differentiating T cells thus causes this immune system-associated disease. These patients are very fragile, requiring simultaneous anti-infective therapy, nutritional support, and immunosuppression. HSCT provides a curative approach.

Functional T Cell Defects

A subset of T cell PIDs with autosomal inheritance is characterized by partially preserved T cell differentiation but defective activation resulting in abnormal effector function (Fig. 344-2). There are many causes of these defects, but all lead to susceptibility to viral and opportunistic infections, chronic diarrhea, and failure to thrive, with onset during childhood. Careful phenotyping and in vitro functional assays are required to identify these diseases, the best characterized of which are the following.

ZETA-ASSOCIATED PROTEIN 70 (ZAP70) DEFICIENCY

Zeta-associated protein 70 (ZAP70) is recruited to the TCR following antigen recognition. A ZAP70 deficiency leads typically to an almost complete absence of CD8+ T cells; CD4+ T cells are present but cannot be activated in vitro by TCR stimulation.

CALCIUM SIGNALING DEFECTS

A small number of patients have been reported who exhibit a profound defect in in vitro T and B cell activation as a result of defective antigen receptor-mediated Ca²⁺ influx. This defect is caused by a mutation in the calcium channel gene (*ORAI1*) or its activator (*STIM-1*). It is noteworthy that these patients are also prone to autoimmune manifestations (blood cytopenias) and exhibit a nonprogressive muscle disease.

HUMAN LEUKOCYTE ANTIGEN (HLA) CLASS II DEFICIENCY

Defective expression of HLA class II molecules is the hallmark of a group of four recessive genetic defects all of which affect molecules (RFX5, RFXAP, RFXANK, and CIITA) involved in the transactivation of the genes coding for HLA class II. As a result, low but variable CD4+ T cell counts are observed in addition to defective antigen-specific T and B cell responses. These patients are particularly susceptible to herpesvirus, adenovirus, and enterovirus infections and chronic gut/liver *Cryptosporidium* infections.

HLA CLASS I DEFICIENCY

Defective expression of molecules involved in antigen presentation by HLA class I molecules (i.e., TAP-1, TAP-2, and Tapasin) leads to reduced CD8+ T cell counts, loss of HLA class I antigen expression, and a particular phenotype consisting of chronic obstructive pulmonary disease and severe vasculitis.

OTHER DEFECTS

A variety of other T cell PIDs have been described, some of which are associated with a precise molecular defect (e.g., IL-2-inducible T cell kinase [ITK] deficiency, IL-21 and IL21 receptor deficiencies, CARD11 deficiency, DOCK2 deficiency, RORC deficiency). These conditions are also characterized by profound vulnerability to infections, such as severe Epstein-Barr virus (EBV)-induced B cell proliferation and autoimmune disorders in ITK deficiency. Milder phenotypes are associated with CD8 and CD3 γ deficiencies.

HSCT is indicated for most of these diseases, although the prognosis is worse than in SCID because many patients are chronically infected at the time of diagnosis. Fairly aggressive immunosuppression and myeloablation may be necessary to achieve engraftment of allogeneic stem cells.

T Cell Primary Immunodeficiencies with DNA Repair Defects

This is a group of PIDs characterized by a combination of T and B cell defects of variable intensity, together with a number of nonimmunologic features resulting from DNA fragility. The autosomal recessive disorder *ataxia-telangiectasia* (AT) is the most frequently encountered condition in this group. It has an incidence of 1:40,000 live births and causes B cell defects (low IgA, IgG2

deficiency, and low antibody production), which often require immunoglobulin replacement. AT is associated with a progressive T cell immunodeficiency. As the name suggests, the hallmark features of AT are telangiectasia and cerebellar ataxia. The latter manifestations may not be detectable before the age of 3–4 years, so that AT should be considered in young children with IgA deficiency and recurrent and problematic infections. Diagnosis is based on a cytogenetic analysis showing excessive chromosomal rearrangements (mostly affecting chromosomes 7 and 14) in lymphocytes. AT is caused by a mutation in the gene encoding the ATM protein—a kinase that plays an important role in the detection and repair of DNA lesions (or cell death if the lesions are too numerous) by triggering several different pathways. Overall, AT is a progressive disease that carries a very high risk of lymphoma, leukemia, and (during adulthood) carcinomas. A variant of AT (“AT-like disease”) is caused by mutation in the *MRE11* gene.

Nijmegen breakage syndrome (NBS) is a less common condition that also results from chromosome instability (with the same cytogenetic abnormalities as in AT). NBS is characterized by a severe T and B cell combined immune deficiency with autosomal recessive inheritance. Individuals with NBS exhibit microcephaly and a bird-like face, but have neither ataxia nor telangiectasia. The risk of malignancies is very high. NBS results from a deficiency in nibrin (NBS1, a protein associated with MRE11 and Rad50 that is involved in checking DNA lesions) caused by hypomorphic mutations.

Severe forms of *dyskeratosis congenita* (also known as Hoyeraal-Hreidarsson syndrome) combine a progressive immunodeficiency that can also include an absence of B and NK lymphocytes, progressive bone marrow failure, microcephaly, in utero growth retardation, and gastrointestinal disease. The disease can be X-linked or, more rarely, autosomal recessive. It is caused by the mutation of genes encoding telomere maintenance proteins, including dyskerin (DKC1).

Finally, *immunodeficiency with centromeric and facial anomalies* (ICF) is a complex syndrome of autosomal recessive inheritance that variably combines a mild T cell immune deficiency with a more severe B cell immune deficiency, coarse face, digestive disease, and mild mental retardation. A diagnostic feature is the detection by cytogenetic analysis of multiradial aspects in multiple chromosomes (most frequently 1, 9, and 16) corresponding to an abnormal DNA structure secondary to defective DNA methylation. It is the consequence of a deficiency in most cases in the DNA methyltransferase DNMT3B, ZBTB24, CDCA7, or HELLS.

T Cell Primary Immunodeficiencies with Hyper-IgE

Several T cell PIDs are associated with elevated serum IgE levels (as in Omenn syndrome). A condition sometimes referred to as *autosomal recessive hyper-IgE syndrome* is notably characterized by recurrent bacterial infections in the skin and respiratory tract and severe skin and mucosal infections by pox viruses and human papillomaviruses, together with severe allergic manifestations. T and B lymphocyte counts are low. Mutations in the *DOCK8* gene have been found in many of these patients. This condition is an indication for HSCT.

A very rare, related condition with autosomal recessive inheritance that causes a similar susceptibility to infection with various microbes (see above), including mycobacteria, reportedly results from a deficiency in Tyk-2, a JAK family kinase involved in the signaling of many different cytokine receptors.

Autosomal Dominant Hyper-IgE Syndrome

This unique condition, the *autosomal dominant hyper-IgE syndrome*, is usually diagnosed by the combination of recurrent skin and lung infections that can be complicated by pneumatoceles. Infections are caused by pyogenic bacteria and fungi. Several other manifestations characterize hyper-IgE syndrome, including facial dysmorphism, defective loss of primary teeth, hyperextensibility, scoliosis, and osteoporosis. Elevated serum IgE levels are typical of this syndrome. Defective T_H17 effector responses have been shown to account at least in part for the specific patterns of susceptibility to particular microbes. This condition is caused by a heterozygous (dominant) mutation in the gene encoding the transcription factor STAT3 that is required in a number of signaling pathways following binding of cytokine to cytokine receptors (such as that of IL-6 and the IL-6 receptor). It also results in partially defective antibody production because of defective IL-21R signaling. Hence, immunoglobulin substitution can be considered as prophylaxis of bacterial infections.

Cartilage Hair Hypoplasia

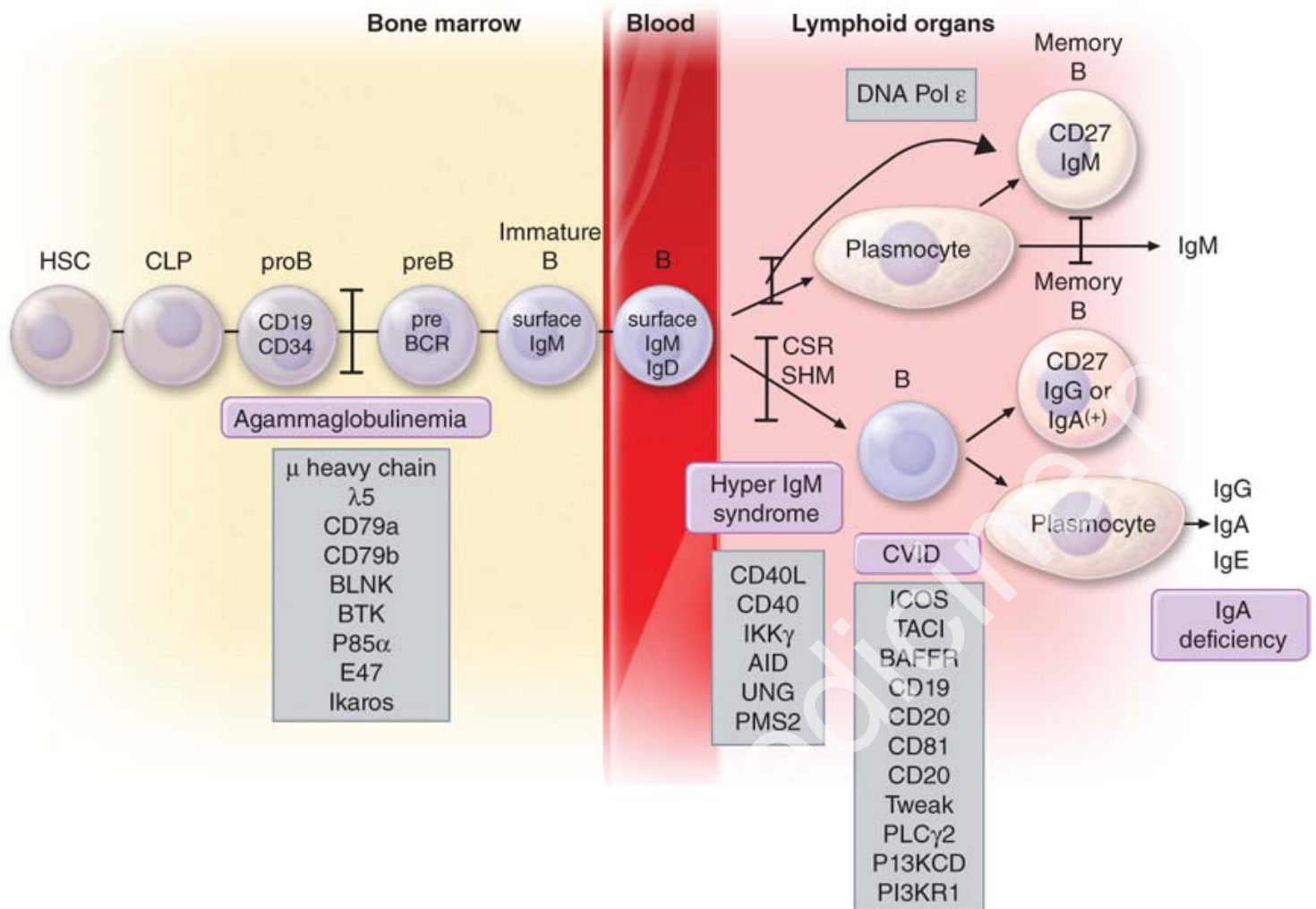
The autosomal recessive *cartilage hair hypoplasia* (CHH) disease is characterized by short-limb dwarfism, metaphyseal dysostosis, and sparse hair, together with a combined T and B cell PID of extremely variable intensity (ranging from quasi-SCID to no clinically significant immune defects). The condition can predispose to erythroblastopenia, autoimmunity, and tumors. It is caused by mutations in the *RMRP* gene for a noncoding ribosome-associated RNA.

CD40 Ligand and CD40 Deficiencies

Hyper-IgM syndrome (HIGM) is a well-known PID that is usually classified as a B cell immune deficiency (see [Fig. 344-4](#) and below). It results from defective immunoglobulin class switch recombination (CSR) in germinal centers and leads to profound deficiency in production of IgG, IgA, and IgE (although IgM production is maintained). Approximately half of HIGM sufferers are also prone to opportunistic infections, for example, interstitial pneumonitis caused by *Pneumocystis jiroveci* (in young children), protracted diarrhea and cholangitis caused by *Cryptosporidium*, and infection of the brain with *Toxoplasma gondii*.

FIGURE 344-4

B cell differentiation and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which give rise to pre-B cells. The B cell differentiation pathway goes through the pre-B cell stage (expression of the μ heavy chain and surrogate light chain), the immature B cell stage (expression of surface IgM), and the mature B cell stage (expression of surface IgM and IgD). The main phenotypic characteristics of these cells are indicated. In lymphoid organs, B cells can differentiate into plasma cells and produce IgM or undergo (in germinal centers) Ig class switch recombination (CSR) and somatic mutation of the variable region of V genes (SHM) that enable selection of high-affinity antibodies. These B cells produce antibodies of various isotypes and generate memory B cells. PIDs are indicated in the purple boxes. CVID, common variable immunodeficiency.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

In the majority of cases, this condition has an X-linked inheritance and is caused by a deficiency in CD40 ligand (L). CD40L induces signaling events in B cells that are necessary for both CSR and adequate activation of other CD40-expressing cells that are involved in innate immune responses against the above-mentioned microorganisms. More rarely, the condition is caused by a deficiency in CD40 itself. The poorer prognosis of CD40L and CD40 deficiencies (relative to most other HIGM conditions) implies that (1) thorough investigations have to be performed in all cases of HIGM and (2) potentially curative HSCT should be discussed on a case-by-case basis for this group of patients.

Wiskott-Aldrich Syndrome

WAS is a complex, recessive, X-linked disease with an incidence of ~1 in 200,000 live births. It is caused by mutations in the *WASP* gene that affect not only T lymphocytes but also the other lymphocyte subsets, dendritic cells, and platelets. WAS is typically characterized by the following clinical manifestations: recurrent bacterial infections, eczema, and bleeding caused by thrombocytopenia. However, these manifestations are highly variable—mostly as a consequence of the many different *WASP* mutations that have been observed. Null mutations predispose affected individuals to invasive and bronchopulmonary infections, viral infections, severe eczema, and autoimmune manifestations. The latter include autoantibody-mediated blood cytopenia, glomerulonephritis, skin and visceral vasculitis (including brain vasculitis), erythema nodosum, and arthritis. Another possible consequence of WAS is lymphoma, which may be virally induced (e.g., by EBV or Kaposi's sarcoma-associated herpesvirus). Thrombocytopenia can be severe and compounded by the peripheral destruction of platelets associated with autoimmune disorders. Hypomorphic mutations usually lead to milder outcomes that are generally limited to thrombocytopenia. It is noteworthy that even patients with "isolated" X-linked thrombocytopenia can develop severe autoimmune disease or lymphoma later in life. The immunologic workup is not very informative; there can be a relative CD8⁺ T cell deficiency, frequently accompanied by low serum IgM levels and decreased antigen-specific antibody responses. A typical

feature is reduced-sized platelets on a blood smear. Diagnosis is based on intracellular immunofluorescence analysis of WAS protein (WASp) expression in blood cells. WASp regulates the actin cytoskeleton and thus plays an important role in many lymphocyte functions, including cell adhesion and migration and the formation of synapses between antigen-presenting and target cells. Predisposition to autoimmune disorders is in part related to defective regulatory T cells. The treatment of WAS should match the severity of disease expression. Prophylactic antibiotics, immunoglobulin G (IgG) supplementation, and careful topical treatment of eczema are indicated. Although splenectomy improves platelet count in a majority of cases, this intervention is associated with a significant risk of infection (both before and after HSCT). Allogeneic HSCT is curative, with fairly good results overall. Gene therapy trials are also under way. A similar condition has been reported in a girl with a deficiency in the Wiskott-Aldrich interacting protein (WIP).

A few other complex PIDs are worth mentioning. *Sp110 deficiency* causes a T cell PID with liver venoocclusive disease and hypogammaglobulinemia. *Chronic mucocutaneous candidiasis* (CMC) is a heterogeneous disease, considering the different inheritance patterns that have been observed. In some cases, chronic candidiasis is associated with late-onset bronchopulmonary infections, bronchiectasis, and brain aneurysms. Moderate forms of CMC are related to autoimmunity and AIRE deficiency (see below). In this setting, predisposition to *Candida* infection is associated with the detection of autoantibodies to T_H17 cytokines. Recently, deficiencies in IL-17A, IL-17F, and IL-17 receptor A and C and in the associated protein Act1, and above all, gain-of-function mutations in *STAT1* have been found to be associated with CMC. In all cases, CMC is related to defective T_H17 function. Innate immunodeficiency in CARD9 also predisposes to chronic invasive fungal infection.

B LYMPHOCYTE DEFICIENCIES

Deficiencies that predominantly affect B lymphocytes are the most frequent PIDs and account for 60–70% of all cases (Table 344-1, Fig. 344-4). B lymphocytes make antibodies. Pentameric IgMs are found in the vascular compartment and are also secreted at mucosal surfaces. IgG antibodies diffuse freely into extravascular spaces, whereas IgA antibodies are produced and secreted predominantly from mucosa-associated lymphoid tissues. Although Ig isotypes have distinct effector functions, including Fc receptor-mediated and (indirectly) C₃ receptor-dependent phagocytosis of microorganisms, they share the ability to recognize and neutralize a given pathogen. Defective antibody production therefore allows the establishment of invasive, pyogenic bacterial infections as well as recurrent sinus and pulmonary infections (mostly caused by *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and, less frequently, gram-negative bacteria). If left untreated, recurrent bronchial infections lead to bronchiectasis and, ultimately, cor pulmonale and death. Parasitic infections such as caused by *Giardia lamblia* and bacterial infections caused by *Helicobacter* and *Campylobacter* of the gut are also observed. A complete lack of antibody production (namely agammaglobulinemia) can also predispose affected individuals to severe, chronic, disseminated enteroviral infections causing meningoencephalitis, hepatitis, and a dermatomyositis-like disease.

Even with the most profound of B cell deficiencies, infections rarely occur before the age of 6 months; this is because of transient protection provided by the transplacental passage of immunoglobulins during the last trimester of pregnancy. Conversely, a genetically nonimmunodeficient child born to a mother with hypogammaglobulinemia is, in the absence of maternal Ig substitution, usually prone to severe bacterial infections in utero and for several months after birth.

Diagnosis of B cell PIDs relies on the determination of serum Ig levels (Table 344-2). Determination of antibody production following immunization with tetanus toxoid vaccine or nonconjugated pneumococcal polysaccharide antigens can also help diagnose more subtle deficiencies. Another useful test is B cell phenotype determination in switched μ - δ - CD27+ and nonswitched memory B cells (μ + δ + CD27+). In agammaglobulinemic patients, examination of bone marrow B cell precursors (Fig. 344-4) can help obtain a precise diagnosis and guide the choice of genetic tests.

Agammaglobulinemia

Agammaglobulinemia is characterized by a profound defect in B cell development (<1% of the normal B cell blood count). In most patients, very low residual Ig isotypes can be detected in the serum. In 85% of cases, agammaglobulinemia is caused by a mutation in the *BTK* gene that is located on the X chromosome. The *BTK* gene product is a kinase that participates in (pre) B cell receptor signaling. When the kinase is defective, there is a block (albeit a leaky one) at the pre-B to B cell stage (Fig. 344-4).

Detection of *BTK* by intracellular immunofluorescence of monocytes, and lack thereof in patients with X-linked agammaglobulinemia (XLA), is a useful diagnostic test. Not all of the mutations in *BTK* result in agammaglobulinemia, since some patients have a milder form of hypogammaglobulinemia and low but detectable B cell counts. These cases should not be confused with common variable immunodeficiency (CVID, see below). About 10% of agammaglobulinemia cases are caused by alterations in genes encoding elements of the pre-B cell receptor, i.e., the μ heavy chain, the $\lambda 5$ surrogate light chain, Ig α or Ig β , the scaffold protein BLNK, and the p85 α subunit of phosphatidylinositol 3 phosphate kinase (PI3K) and the Ikaros transcription factor. In 5% of cases, the defect is unknown. It is noteworthy that agammaglobulinemia can be observed in patients with ICF syndrome, despite the presence of normal peripheral B cell counts. Lastly, agammaglobulinemia can be a manifestation of a myelodysplastic syndrome (associated or not with neutropenia). Treatment of agammaglobulinemic patients is based on immunoglobulin replacement (see below). Profound hypogammaglobulinemia is also observed in adults, in association with thymoma.

Hyper-IgM (HIGM) Syndromes

HIGM is a rare B cell PID characterized by defective Ig CSR. It results in very low serum levels of IgG and IgA and elevated or normal serum IgM levels. The clinical severity is similar to that seen in agammaglobulinemia, although chronic lung disease and sinusitis are less frequent and enteroviral infections are uncommon. As discussed above, a diagnosis of HIGM involves screening for an X-linked CD40L deficiency and an autosomal recessive CD40 deficiency, which affect both B and T cells. In 50% of cases affecting only B cells, these isolated HIGM syndromes result from mutations in the gene encoding activation-induced deaminase, the protein that induces CSR in B cell germinal centers. These patients usually have enlarged lymphoid organs. In the other 50% of cases, the etiology is unknown (except for rare UNG and PMS2 deficiencies). Furthermore, IgM-mediated autoimmunity and lymphomas can occur in HIGM syndrome. It is noteworthy that HIGM can result from fetal rubella syndrome or can be a predominant immunologic feature of other PIDs, such as the immunodeficiency associated with ectodermic anhydrotic hypoplasia X-linked NEMO deficiency and the combined T and B cell PIDs caused by DNA repair defects such as AT and Cernunnos deficiency.

Common Variable Immunodeficiency

CVID is an ill-defined condition characterized by low serum levels of one or more Ig isotypes. Its prevalence is estimated to be 1 in 20,000. The condition is recognized predominantly in adults, although clinical manifestations can occur earlier in life. Hypogammaglobulinemia is associated with at least partially defective antibody production in response to vaccine antigens. B lymphocyte counts are often normal but can be low. Besides infections, CVID patients may develop lymphoproliferation (splenomegaly), granulomatous lesions, colitis, antibody-mediated autoimmune disease, and lymphomas. A family history is found in 10% of cases. A clear-cut dominant inheritance pattern is found in some families, whereas recessive inheritance is observed more rarely. In most cases, no molecular cause can be identified. A small number of patients in Germany were found to carry mutations in the *ICOS* gene encoding a T cell-membrane protein that contributes to B cell activation and survival. In 10% of patients with CVID, monoallelic or biallelic mutations of the gene encoding TACI (a member of the tumor necrosis factor [TNF] receptor family that is expressed on B cells) have been found. In fact, heterozygous TACI mutations correspond to a genetic susceptibility factor, since similar heterozygous mutations are found in 1% of controls. The B-cell activating factor (BAFF) receptor was found to be defective in a kindred with CVID, although not all individuals carrying the mutation have CVID. Recently a group of patients with hypogammaglobulinemia and lymphoproliferation was shown to exhibit dominant gain of function mutations in the *PIK3CD* gene encoding the p110 δ form of PI3 kinase or in the *PI3KR1* gene encoding the regulatory p85 α subunit of PI3 kinase. Rare cases of hypogammaglobulinemia were found to be associated with CD19 and CD81 deficiencies. These patients have B cells that can be identified by typing for other B cell markers.

A diagnosis of CVID should be made after excluding the presence of hypomorphic mutations associated with agammaglobulinemia or more subtle T cell defects; this is particularly the case in children. It is possible that many cases of CVID result from a constellation of factors, rather than a single genetic defect. Hypogammaglobulinemia can be associated with neutropenia and lymphopenia in the warts, hypogammaglobulinemia, infections, and myelokathexis syndrome (WHIM) caused by dominant gain-of-function mutation of *CXCR4*, resulting in cell retention in the bone marrow.

Selective Ig Isotype Deficiencies

IgA deficiency and CVID represent polar ends of a clinical spectrum due to the same underlying gene defect(s) in a large subset of these patients. IgA deficiency is the most common PID; it can be found in 1 in every 600 individuals. It is asymptomatic in most cases; however, individuals may present with increased numbers of acute and chronic respiratory infections that may lead to bronchiectasis. In addition, over their lifetime, these patients experience an increased susceptibility to drug allergies, atopic disorders, and autoimmune diseases. Symptomatic IgA deficiency is probably related to CVID, since it can be found in relatives of patients with CVID. Furthermore, IgA deficiency may progress to CVID. It is thus important to assess serum Ig levels in IgA-deficient patients (especially when infections occur frequently) in order to detect changes that should prompt the initiation of immunoglobulin replacement. Selective IgG2 (+G4) deficiency (which in some cases may be associated with IgA deficiency) can also result in recurrent sinopulmonary infections and should thus be specifically sought in this clinical setting. These conditions are ill-defined and often transient during childhood. A pathophysiologic explanation has not been found.

Selective Antibody Deficiency to Polysaccharide Antigens

Some patients with normal serum Ig levels are prone to *S. pneumoniae* and *H. influenzae* infections of the respiratory tract. Defective production of antibodies against polysaccharide antigens (such as those in the *S. pneumoniae* cell wall) can be observed and is probably causative. This condition may correspond to a defect in marginal zone B cells, a B cell subpopulation involved in T-independent antibody responses.

Immunoglobulin Replacement

IgG antibodies have a half-life of 21–28 days. Thus, injection of plasma-derived polyclonal IgG containing a myriad of high-affinity antibodies can provide protection against disease-causing microorganisms in patients with defective IgG antibody production. This form of therapy should not be based on laboratory data alone (i.e., IgG and/or antibody deficiency) but should be guided by the occurrence or not of infections; otherwise, patients might be subjected to unjustified IgG infusions. Immunoglobulin replacement can be performed by IV or subcutaneous routes. In the former case, injections have to be repeated every 3–4 weeks, with a residual target level of 800 mg/mL in patients who had very low IgG levels prior to therapy. Subcutaneous injections are typically performed once a week, although the frequency can be adjusted on a case-by-case basis. A trough level of 800 mg/mL is desirable. Whatever the mode of administration, the main goal is to reduce the frequency of the respiratory tract infections and prevent chronic lung and sinus disease. The two routes appear to be equally safe and efficacious, and so the choice should be left to the preference of the patient.

In patients with chronic lung disease, chest physical therapy with good pulmonary toilet and the cyclic use of antibiotics are also needed. Immunoglobulin replacement is well tolerated by most patients, although the selection of the best-tolerated Ig preparation may be necessary in certain cases. Since IgG preparations contain a small proportion of IgAs, caution should be taken in patients with residual antibody production capacity and a complete IgA deficiency, as these subjects may develop anti-IgA antibodies that can trigger anaphylactic shock. These patients should be treated with IgA-free IgG preparations. Immunoglobulin replacement is a lifelong therapy; its rationale and procedures have to be fully understood and mastered by the patient and his or her family in order to guarantee the strict observance required for efficacy.

PRIMARY IMMUNODEFICIENCIES AFFECTING REGULATORY PATHWAYS

An increasing number of PIDs have been found to cause homeostatic dysregulation of the immune system, either alone or in association with increased vulnerability to infections (Table 344-1). Defects of this type affecting the innate immune system and autoinflammatory syndromes will not be covered in this chapter. However, three specific entities (hemophagocytic lymphohistiocytosis [HLH], lymphoproliferation, and autoimmunity) will be described below.

HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

HLH is characterized by an unremitting activation of CD8+ T lymphocytes and macrophages that leads to organ damage (notably in the liver, bone marrow, and central nervous system). This syndrome results from a broad set of inherited diseases, most of which impair T and NK lymphocyte cytotoxicity. The manifestations of HLH are often induced by a viral infection. EBV is the most frequent trigger. In severe forms of HLH, disease onset may start during the first year of life or even (in rare cases) at birth.

Diagnosis relies on the identification of the characteristic symptoms of HLH (fever, hepatosplenomegaly, edema, neurologic diseases, blood cytopenia, increased liver enzymes, hypofibrinogenemia, high triglyceride levels, elevated markers of T cell activation, and hemophagocytic features in the bone marrow or cerebrospinal fluid). Functional assays of postactivation cytotoxic granule exocytosis (CD107 fluorescence at the cell membrane) can suggest genetically determined HLH. The conditions can be classified into three subsets:

1. Familial HLH with autosomal recessive inheritance, including perforin deficiency (30% of cases) that can be recognized by assessing intracellular perforin expression; Munc13-4 deficiency (30% of cases); syntaxin 11 deficiency (10% of cases); Munc18-2 deficiency (20% of cases); and a few residual cases that lack a known molecular defect.
2. HLH with partial albinism. Three conditions combine HLH and abnormal pigmentation, where hair examination can help in the diagnosis: Chédiak-Higashi syndrome, Griscelli syndrome, and Hermansky-Pudlak syndrome type II. Chédiak-Higashi syndrome is also characterized by the presence of giant lysosomes within leukocytes (**Chap. 60**), in addition to a primary neurologic disorder with slow progression of symptoms over time.
3. XLP is characterized in most patients by the induction of HLH following EBV infection, while other patients develop progressive hypogammaglobulinemia similar to what is observed in CVID and/or certain lymphomas. XLP is caused by a mutation in the *SH2D1A* gene that encodes the adaptor protein SAP (associated with a SLAM family receptor). Several immunologic abnormalities have been described, including low 2B4-mediated NK cell cytotoxicity, impaired differentiation of NKT cells, defective antigen-induced T cell death, and defective T cell helper activity for B cells. A related disorder (XLP2) has recently been described. It is also X-linked and induces HLH (frequently after EBV infection), although the clinical manifestation may be less pronounced. The condition is associated with a deficiency of the antiapoptotic molecule XIAP. The pathophysiology of XLP2 remains unclear; however, it may be related to control of inflammation in macrophages as there is a functional link between XIAP and NLRC4, an inflammasome component, in which gain of function can also induce HLH. XLP2 is also frequently associated with colitis.

HLH is a life-threatening complication. The treatment of this condition requires aggressive immunosuppression with either the cytotoxic agent etoposide or anti-T cell antibodies; specific therapy targeting interferon γ , which is critical in causing HLH, is an additional option to consider. Once remission has been achieved, HSCT should be performed, since it provides the only curative form of therapy.

AUTOIMMUNE LYMPHOPROLIFERATIVE SYNDROME

Autoimmune lymphoproliferative syndrome (ALPS) is characterized by nonmalignant T and B lymphoproliferation causing splenomegaly and enlarged lymph nodes; 70% of patients also display autoimmune manifestations such as autoimmune cytopenias, Guillain-Barré syndrome, uveitis, and hepatitis (**Chaps. 62 and 342**). A hallmark of ALPS is the presence of CD4–CD8–TCR $\alpha\beta$ + T cells (2–50%) in the blood of affected individuals. Hypergammaglobulinemia involving IgG and IgA is also frequently observed. The syndrome is caused by a defect in Fas-mediated apoptosis of lymphocytes, which can thus accumulate and mediate autoimmunity. Furthermore, ALPS can lead to malignancies.

Most patients carry a heterozygous mutation in the gene encoding Fas that is characterized by dominant inheritance and variable penetrance, depending on the nature of the mutation. A rare and severe form of the disease with early onset can be observed in patients carrying a biallelic mutation of Fas, which profoundly impairs the protein's expression and/or function. Fas-ligand, caspase 10, caspase 8, and somatic neuroblastoma RAS viral oncogene homologue (NRAS) mutations have also been reported in a few cases of ALPS. Many cases of ALPS have not been precisely delineated at the molecular level. A B cell–predominant ALPS has recently been found associated with a protein kinase C δ gene mutation. Treatment of ALPS is essentially based on the use of proapoptotic drugs, which need to be carefully administered in order to avoid toxicity.

COLITIS, AUTOIMMUNITY, AND PRIMARY IMMUNODEFICIENCIES

Several PIDs (most of which are T cell–related) can cause severe gut inflammation. The prototypic example is *immunodysregulation polyendocrinopathy enteropathy X-linked syndrome* (IPEX), characterized by a widespread inflammatory enteropathy, food intolerance, skin rashes, autoimmune cytopenias, and diabetes. The syndrome is caused by loss-of-function

mutations in the gene encoding the transcription factor FOXP3, which is required for the acquisition of effector function by regulatory T cells. In most cases of IPEX, CD4+CD25+ regulatory T cells are absent from the blood. This condition has a poor prognosis and requires aggressive immunosuppression. The only possible curative approach is allogeneic HSCT. IPEX-like syndromes that lack a FOXP3 mutation have also been described. In some cases, a CD25 deficiency has been found. Defective CD25 expression also impairs regulatory cell expansion/function. This functional T cell deficiency means that CD25-deficient patients are also at increased risk of opportunistic infections. It is noteworthy that abnormalities in regulatory T cells have been described in other PID settings, such as in Omenn syndrome, STAT5b deficiency, STIM1 (Ca flux) deficiency, and WAS; these abnormalities may account (at least in part) for the occurrence of inflammation and autoimmunity. The autoimmune features observed in a small fraction of patients with DiGeorge's syndrome may have the same cause. Recently, severe inflammatory gut disease has been described in patients with a deficiency in the IL-10 receptor or IL-10.

Dominant mutations in genes encoding the regulatory molecule CTLA4, recessive mutations in the gene encoding LRBA (a molecule involved in recycling of CTLA4) as well as dominant gain of function mutation of STAT3 cause a multifaceted lymphoproliferative and autoimmune syndrome, frequently involving inflammatory bowel disease that can be associated with hypogammaglobulinemia. Molecular diagnosis is required before adapted targeted therapies are undertaken.

A distinct autoimmune entity is observed in *autoimmune polyendocrinopathy candidiasis ectodermal dysplasia* (APECED) syndrome, which is characterized by autosomal recessive inheritance. It consists of multiple autoimmune manifestations that can affect solid organs in general and endocrine glands in particular. Mild, chronic *Candida* infection is often associated with this syndrome. The condition is due to mutations in the autoimmune regulator (*AIRE*) gene and results in impaired thymic expression of self-antigens by medullary epithelial cells and impaired negative selection of self-reactive T cells that leads to autoimmune manifestations.

A combination of hypogammaglobulinemia, autoantibody production, cold-induced urticaria or skin granulomas, or autoinflammation has been reported, and has been termed the *PLCY2-associated antibody deficiency and immune dysregulation* (PLAID or APLAID).

CONCLUSION

The variety and complexity of the clinical manifestations of the many different PIDs strongly indicate that it is important to raise awareness of these diseases. Indeed, early diagnosis is essential for establishing an appropriate therapeutic regimen. Hence, patients with suspected PIDs must always be referred to experienced clinical centers that are able to perform appropriate molecular and genetic tests. A precise molecular diagnosis is not only necessary for initiating the most suitable treatment, but is also important for genetic counseling and prenatal diagnosis.

One pitfall that may hamper diagnosis is the high variability that is associated with many PIDs. Variable disease expression can result from the differing consequences of various mutations associated with a given condition, as exemplified by WAS and, to a lesser extent, XLA. There can also be effects of modifier genes (as also suspected in XLA) and environmental factors such as EBV infection that can be the main trigger of disease in XLP conditions. Furthermore, it has recently been established that somatic mutations in an affected gene can attenuate the phenotype of a number of T cell PIDs. This has been described for ADA deficiency, X-linked SCID, RAG deficiencies, NF- κ B essential modulator (NEMO) deficiency, and, most frequently, WAS. In contrast, somatic mutations can create disease states analogous to PID, as reported for ALPS. Lastly, cytokine-neutralizing autoantibodies can mimic a PID, as shown for IFN- γ .

Many aspects of the pathophysiology of PIDs are still unknown, and the disease-causing gene mutations have not been identified in all cases (as illustrated by CVID and IgA deficiency). However, our medical understanding of PIDs has now reached the stage where scientifically based approaches to the diagnosis and treatment of these diseases can be implemented. A genetic diagnosis has become a milestone step in the care of PID patients.

FURTHER READING

Casanova JL et al: Guidelines for genetic studies in single patients: Lessons from primary immunodeficiencies. *J Exp Med* 211:2137, 2014. [[PubMed: 25311508](#)]

Fischer A et al: Gene therapy for primary immunodeficiencies. *Clin Genet* 88:507, 2015. [[PubMed: 25708106](#)]

Fischer A et al: Severe combined immunodeficiencies and related disorders. *Nat Rev Dis Primers* 1:15061, 2015. [[PubMed: 27189259](#)]

Holland SM: Chronic granulomatous disease. *Hematol Oncol Clin North Am* 2013;27:89, viii. [[PubMed: 23351990](#)]

Kwan A et al: Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA* 312:729, 2014. [[PubMed: 25138334](#)]

Notarangelo LD: Functional T cell immunodeficiencies (with T cells present). *Annu Rev Immunol* 31:195, 2013. [[PubMed: 23298211](#)]

Ochs HD et al (eds): *Primary Immunodeficiencies: A Molecular and Genetic Approach*. New York, Oxford University Press, 2013.

Picard C et al: International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. *J Clin Immunol* 38:96, 2018.

Picard C, Fischer A: Contribution of high-throughput DNA sequencing to the study of primary immunodeficiencies. *Eur J Immunol* 44:2854, 2014. [[PubMed: 25154746](#)]

Resnick ES, Cunningham-Rundles C: The many faces of the clinical picture of common variable immune deficiency. *Curr Opin Allergy Clin Immunol* 12:595, 2012. [[PubMed: 23026770](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 345: Urticaria, Angioedema, and Allergic Rhinitis

Katherine N. Cahill; Joshua A. Boyce

INTRODUCTION

The term *atopy* implies a tendency to manifest asthma, rhinitis, urticaria, and atopic dermatitis alone or in combination, in association with the presence of allergen-specific IgE. However, individuals without an atopic background may also develop hypersensitivity reactions, particularly urticaria and anaphylaxis, associated with the presence of IgE. Since mast cells are key effector cells in allergic rhinitis and asthma, and the dominant effector in urticaria, anaphylaxis, and systemic mastocytosis, its developmental biology, activation pathway, product profile, and target tissues will be considered in the introduction to these clinical disorders. Dysregulation of mast cell development seen in mastocytosis will be covered in a separate chapter.

The binding of IgE to human mast cells and basophils, a process termed *sensitization*, prepares these cells for subsequent antigen-specific activation. The high-affinity Fc receptor for IgE, designated FcεRI, is composed of one α, one β, and two disulfide-linked γ chains, which together cross the plasma membrane seven times. The α chain is responsible for IgE binding, and the β and γ chains provide for signal transduction that follows the aggregation of the sensitized tetrameric receptors by polymeric antigen. The binding of IgE stabilizes the α chain at the plasma membrane, thus increasing the density of FcεRI receptors at the cell surface while sensitizing the cell for effector responses. This accounts for the correlation between serum IgE levels and the numbers of FcεRI receptors detected on circulating basophils. Signal transduction is initiated through the action of a Src family-related tyrosine kinase termed Lyn that is constitutively associated with the β chain. Lyn transphosphorylates the canonical immunoreceptor tyrosine-based activation motifs (ITAMs) of the β and γ chains of the receptor, resulting in recruitment of more active Lyn to the β chain and of Syk tyrosine kinase. The phosphorylated tyrosines in the ITAMs function as binding sites for the tandem *src* homology two (SH2) domains within Syk. Syk activates not only phospholipase Cγ, which associates with the linker of activated T cells at the plasma membrane, but also phosphatidylinositol 3-kinase to provide phosphatidylinositol-3,4,5-trisphosphate, which allows membrane targeting of the Tec family kinase Btk and its activation by Lyn. In addition, the Src family tyrosine kinase Fyn becomes activated after aggregation of IgE receptors and phosphorylates the adapter protein Gab2 that enhances activation of phosphatidylinositol 3-kinase. Indeed, this additional input is essential for mast cell activation, but it can be partially inhibited by Lyn, indicating that the extent of mast cell activation is in part regulated by the interplay between these Src family kinases. Activated phospholipase Cγ cleaves phospholipid membrane substrates to provide inositol-1,4,5-trisphosphate (IP₃) and 1,2-diacylglycerols (1,2-DAGs) so as to mobilize intracellular calcium and activate protein kinase C, respectively. The subsequent opening of calcium-regulated activated channels provides the sustained elevations of intracellular calcium required to recruit the mitogen-activated protein kinases, ERK, JNK, and p38 (serine/threonine kinases), which provide cascades to augment arachidonic acid release and to mediate nuclear translocation of transcription factors for various cytokines. The calcium ion-dependent activation of phospholipases cleaves membrane phospholipids to generate lysophospholipids, which, like 1,2-DAG, may facilitate the fusion of the secretory granule perigranular membrane with the cell membrane, a step that releases the membrane-free granules containing the preformed mast cell mediators.

The secretory granule of the human mast cell has a crystalline structure, unlike mast cells of lower species. IgE-dependent cell activation results in solubilization and swelling of the granule contents within the first minute of receptor perturbation; this reaction is followed by the ordering of intermediate filaments about the swollen granule, movement of the granule toward the cell surface, and fusion of the perigranular membrane with that of other granules and with the plasmalemma to form extracellular channels for mediator release while maintaining cell viability.

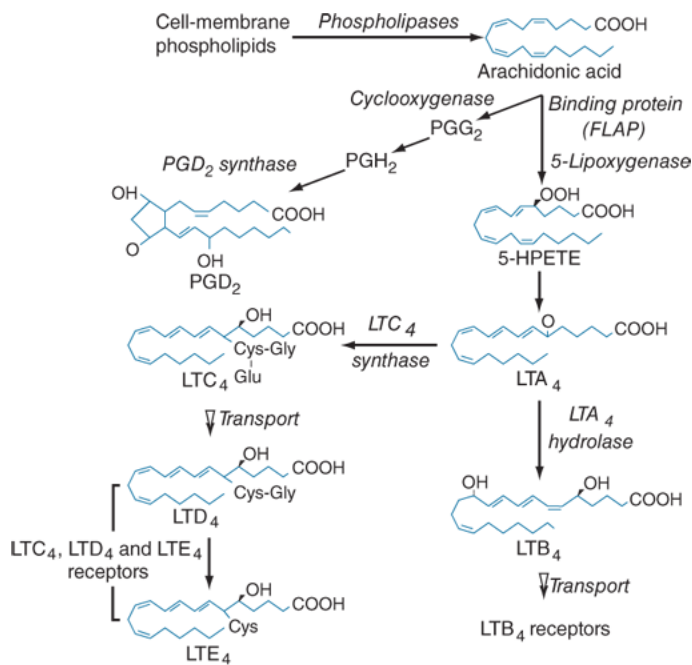
In addition to exocytosis, aggregation of FcεRI initiates two other pathways for generation of bioactive products, namely, lipid mediators and cytokines. The biochemical steps involved in expression of such cytokines as tumor necrosis factor α (TNF-α), interleukin (IL) 1, IL-6, IL-4, IL-5, IL-13, granulocyte-macrophage colony-stimulating factor (GM-CSF), and others, including an array of chemokines, have not been specifically defined for mast cells. Inhibition studies of cytokine production (IL-1β, TNF-α, and IL-6) in mouse mast cells with [cyclosporine](#) or FK506 reveal binding to the ligand-specific immunophilin and attenuation of the calcium ion- and calmodulin-dependent serine/threonine phosphatase, calcineurin.

Lipid mediator generation ([Fig. 345-1](#)) involves translocation of calcium ion-dependent cytosolic phospholipase A₂ to the outer nuclear membrane, with subsequent release of arachidonic acid for metabolic processing by the distinct prostanoid and leukotriene pathways. The constitutive prostaglandin endoperoxide synthase-1 (PGHS-1/cyclooxygenase-1) and the de novo inducible PGHS-2 (cyclooxygenase-2) convert released arachidonic acid to the sequential intermediates, prostaglandins G₂ and H₂. The glutathione-dependent hematopoietic prostaglandin D₂ (PGD₂) synthase then converts PGH₂ to PGD₂, the predominant mast cell prostanoid. The PGD₂ receptor DP₁ is expressed by platelets, natural killer cells, dendritic cells, and epithelial cells, whereas DP₂ is expressed by T_H2 lymphocytes, innate lymphoid type 2 cells, eosinophils, and basophils. Mast cells also generate thromboxane A₂ (TXA₂), a short lived but powerful mediator that induces bronchoconstriction and platelet activation through the T prostanoid (TP) receptor.

FIGURE 345-1

Pathways for biosynthesis and release of membrane-derived lipid mediators from mast cells. In the 5-lipoxygenase pathway, leukotriene A₄ (LTA₄) is the intermediate from which the terminal-pathway enzymes generate the distinct final products, leukotriene C₄ (LTC₄) and leukotriene B₄ (LTB₄), which leave the cell by separate

saturable transport systems. Gamma glutamyl transpeptidase and a dipeptidase then cleave glutamic acid and glycine from LTC₄ to form LTD₄ and LTE₄, respectively. The major mast cell product of the cyclooxygenase system is PGD₂.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

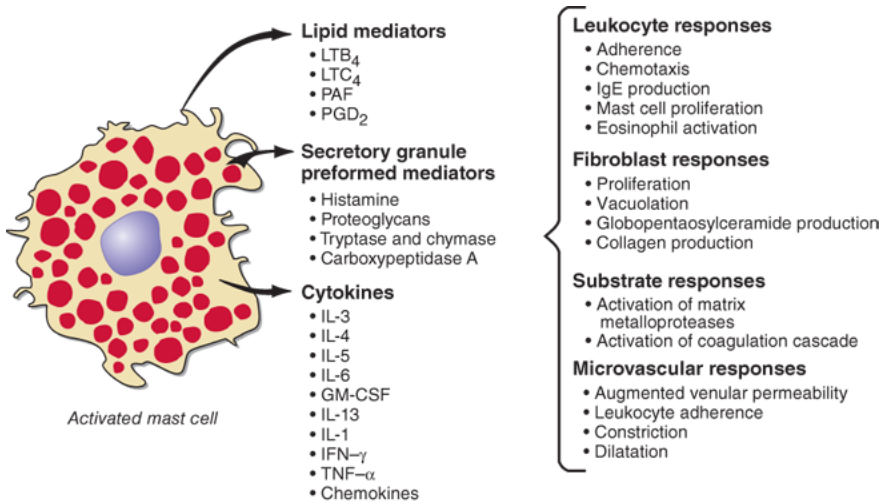
For leukotriene biosynthesis, the released arachidonic acid is metabolized by 5-lipoxygenase (5-LO) in the presence of an integral nuclear membrane protein, 5-LO activating protein (FLAP). The calcium ion-dependent translocation of 5-LO to the nuclear membrane converts the arachidonic acid to the sequential intermediates, 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and leukotriene (LT) A₄. LTA₄ is conjugated with reduced glutathione by LTC₄ synthase, an integral nuclear membrane protein homologous to FLAP. Intracellular LTC₄ is released by a carrier-specific export step for extracellular metabolism to the additional cysteinyl leukotrienes, LTD₄ and LTE₄, by the sequential removal of glutamic acid and glycine. Alternatively, cytosolic LTA₄ hydrolase converts some LTA₄ to the dihydroxy leukotriene LTB₄, which also undergoes specific export. Two receptors for LTB₄, BLT₁ and BLT₂, mediate chemotaxis of human neutrophils. Two receptors for the cysteinyl leukotrienes, CysLT₁ and CysLT₂, are present on smooth muscle of the airways and the microvasculature and on hematopoietic cells such as macrophages, eosinophils, and mast cells. Whereas the CysLT₁ receptor has a preference for LTD₄ and is blocked by the receptor antagonists in clinical use, the CysLT₂ receptor is equally responsive to LTD₄ and LTC₄, is unaffected by these antagonists, and is a negative regulator of the function of the CysLT₁ receptor. LTD₄, acting at CysLT₁ receptors, is the most potent known bronchoconstrictor, whereas LTE₄ induces a vascular leak and mediates the recruitment of eosinophils to the bronchial mucosa. Recently, GPR99, CysLT3 receptor, was identified as an LTE₄ receptor. The lysophospholipid formed during the release of arachidonic acid from 1-O-alkyl-2-acyl-*sn*-glycyl-3-phosphorylcholine can be acetylated in the second position to form platelet-activating factor (PAF). Serum levels of PAF correlated positively with the severity of anaphylaxis to peanut in a recent study, whereas the levels of PAF acetyl hydrolase (a PAF-degrading enzyme) were inversely related to the same outcome.

Unlike most other cells of bone marrow origin, mast cells circulate as committed progenitors lacking their characteristic secretory granules. These committed progenitors express *c-kit*, the receptor for stem cell factor (SCF). Unlike most other lineages, they retain and increase *c-kit* expression with maturation. The SCF interaction with *c-kit* is an absolute requirement for the development of constitutive tissue mast cells residing in skin and connective tissue sites and for the accumulation of mast cells at mucosal surfaces during T_H2-type immune responses. Several T cell-derived cytokines (IL-3, IL-4, IL-5, and IL-9) can potentiate SCF-dependent mast cell proliferation and/or survival in vitro in mice and humans. Indeed, mast cells are absent from the intestinal mucosa in clinical T cell deficiencies, but are present in the submucosa. Based on the immunodetection of secretory granule neutral proteases, mast cells in the lung parenchyma and intestinal mucosa selectively express tryptase, and those in the intestinal and airway submucosa, perivascular spaces, skin, lymph nodes, and breast parenchyma express tryptase, chymase, and carboxypeptidase A (CPA). In the mucosal epithelium of severe asthmatics, mast cells can express tryptase and CPA without chymase. The secretory granules of mast cells selectively positive for tryptase exhibit closed scrolls with a periodicity suggestive of a crystalline structure by electron microscopy, whereas the secretory granules of mast cells with multiple proteases are scroll-poor, with an amorphous or lattice-like appearance.

Mast cells are distributed at cutaneous and mucosal surfaces and in submucosal tissues about venules and could influence the entry of foreign substances by their rapid response capability (Fig. 345-2). Upon stimulus-specific activation and secretory granule exocytosis, histamine and acid hydrolases are solubilized, whereas the neutral proteases, which are cationic, remain largely bound to the anionic proteoglycans, heparin and chondroitin sulfate E, with which they function as a complex. Histamine and the various lipid mediators (PGD₂, LTC₄/D₄/E₄, PAF) alter venular permeability, thereby allowing influx of plasma proteins such as complement and immunoglobulins, whereas LTB₄ mediates leukocyte-endothelial cell adhesion and subsequent directed migration (chemotaxis). The accumulation of leukocytes and plasma opsonins facilitates defense of the microenvironment. The inflammatory response can also be detrimental, as in asthma, where the smooth-muscle constrictor activity of the cysteinyl leukotrienes is evident and much more potent than that of histamine.

FIGURE 345-2

Bioactive mediators of three categories generated by IgE-dependent activation of murine mast cells can elicit common but sequential target cell effects leading to acute and sustained inflammatory responses. GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin; IFN, interferon; LT, leukotriene; PAF, platelet-activating factor; PGD₂, prostaglandin D₂; TNF, tumor necrosis factor.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The cellular component of the mast cell-mediated inflammatory response is augmented and sustained by cytokines and chemokines. IgE-dependent activation of human skin mast cells in situ elicits TNF- α production and release, which in turn induces endothelial cell responses favoring leukocyte adhesion. Similarly, activation of purified human lung mast cells or cord blood-derived cultured mast cells in vitro results in substantial production of proinflammatory (TNF- α) and immunomodulatory cytokines (IL-4, IL-5, IL-13) and chemokines. Bronchial biopsy specimens from patients with asthma reveal that mast cells are immunohistochemically positive for IL-4 and IL-5, but that the predominant localization of IL-4, IL-5, and GM-CSF is to T cells, defined as T_H2 by this profile. IL-4 modulates the T cell phenotype to the T_H2 subtype, determines the isotype switch to IgE (as does IL-13), and upregulates Fc ϵ R1-mediated expression of cytokines by mast cells based on in vitro studies.

An immediate and late cellular phase of allergic inflammation can be induced in the skin, nose, or lung of some allergic humans with local allergen challenge. The immediate phase in the nose involves pruritus and watery discharge; in the lung, it involves bronchospasm and mucus secretion; and in the skin, it involves a wheal-and-flare response with pruritus. The reduced nasal patency, reduced pulmonary function, or erythema with swelling at the skin site in a late-phase response at 6–8 h is associated with biopsy findings of infiltrating and activated T_H2 cells, eosinophils, basophils, and some neutrophils. The progression from early mast cell activation to late cellular infiltration has been used as an experimental surrogate of rhinitis or asthma. However, in asthma, there is an intrinsic hyperreactivity of the airways independent of the associated inflammation. Moreover, early- and late-phase responses (at least in the lung) are far more sensitive to blockade of IgE-dependent mast cell activation (or actions of histamine and cysteinyl leukotrienes) than are spontaneous or virally induced asthma exacerbations.

Consideration of the mechanism of immediate-type hypersensitivity diseases in the human has focused largely on the IgE-dependent recognition of otherwise innocuous substances. A region of chromosome 5 (5q23–31) contains genes implicated in the control of IgE levels including IL-4 and IL-13, as well as IL-3 and IL-9, which are involved in mucosal mast cell hyperplasia, and IL-5 and GM-CSF, which are central to eosinophil development and their enhanced tissue viability. Genes with linkage to the specific IgE response to particular allergens include those encoding the major histocompatibility complex (MHC) and certain chains of the T cell receptor (TCR- $\alpha\delta$). The complexity of atopy and the associated diseases includes susceptibility, severity, and therapeutic responses, each of which is among the separate variables modulated by both innate and adaptive immune stimuli.

The induction of allergic disease requires sensitization of a predisposed individual to a specific allergen. The greatest propensity for the development of atopic allergy occurs in childhood and early adolescence. The allergen is processed by antigen-presenting cells of the monocytic lineage (particularly dendritic cells) located throughout the body at surfaces that contact the outside environment, such as the nose, lungs, eyes, skin, and intestine. These antigen-presenting cells present the epitope-bearing peptides via their MHC to T helper cells and their subsets. The T cell response depends both on cognate recognition and on the cytokine microenvironment provided by the antigen-presenting dendritic cells, with IL-4 directing a T_H2 subset, interferon (IFN) γ a T_H1 profile, and IL-6 with transforming growth factor β (TGF- β) a T_H17 subset. Allergens not only present antigenic epitopes via dendritic cells but also contain pattern recognition ligands that facilitate the immune response by direct initiation of cytokine generation from innate cell types such as basophils, mast cells, eosinophils, and others. The T_H2 response is associated with activation of specific B cells that can also present allergens or that transform into plasma cells for antibody production. Synthesis and release into the plasma of allergen-specific IgE results in sensitization of Fc ϵ R1-bearing cells such as mast cells and basophils, which become activated on exposure to the specific allergen. In certain diseases, including those associated with atopy, the monocyte and eosinophil populations can express a trimeric Fc ϵ R1, which lacks the β chain, and yet respond to its aggregation. An additional recently recognized class of *c-kit*-expressing innate cells (termed group 2 innate lymphoid cells of ILC2) can generate large quantities of IL-5 and IL-13 during antihelminth responses, are prominent in nasal polyps from humans, and contribute to inflammation in allergic diseases.

URTICARIA AND ANGIOEDEMA

DEFINITION

Urticaria and angioedema represent the same pathophysiologic process occurring at different levels of the skin. Urticaria involves dilation of vascular structures in the superficial dermis, while angioedema originates from the deeper dermis and subcutaneous tissues. Not surprisingly they often appear together, with roughly 40% of patients reporting both, and affect >20% of the population at sometime during their lifespan. Urticaria can occur on any area of the body as well-circumscribed wheals with erythematous raised serpiginous borders and blanched centers that may coalesce to become giant wheals. Urticarial lesions last for <24 h, frequently migrate around the body, leave no bruising or scarring and are intensely pruritic. Angioedema is marked by dramatic swelling with more pain than pruritus and minimal erythema, which may develop with a pruritic prodrome and takes hours to days to resolve. Acute urticaria and/or angioedema are episodes that occur for <6 weeks' duration, whereas attacks persisting >6 weeks are designated chronic.

PREDISPOSING FACTORS AND ETIOLOGY

Acute or chronic urticaria and/or angioedema can occur at any point in the lifespan with the third to fifth decade the most common for chronic. Women are affected more often than men with a slight predominance for those with a history of atopy. Acute urticaria is most often the result of exposure to a food, environmental or drug allergen or viral infection while chronic urticaria is often idiopathic.

The classification of urticaria-angioedema presented in [Table 345-1](#) focuses on the different mechanisms for eliciting clinical disease and can be useful for differential diagnosis.

TABLE 345-1

Classification of Urticaria and/or Angioedema

ACUTE	CHRONIC
Drug Reactions NSAIDS, IV contrast, angiotensin-converting enzyme (ACE) inhibitors, etc. Foods Inhalation or contact with environmental allergens Transfusion reactions Insects Infections—viral, bacterial, parasitic	Idiopathic—a subset with autoimmune component Collagen vascular disease—urticarial vasculitis Physical stimuli Dermatographism Cholinergic urticaria Vibration, cold, pressure, water (aquagenic) Sun (solar) Mastocytosis/Urticaria pigmentosa Hereditary Hereditary angioedema (HAE) Familial cold urticaria C3b inhibitor deficiency Muckle-Wells syndrome Schnitzler syndrome Hypereosinophilic syndrome Gleich syndrome

Additional etiologies include physical stimuli such as cold, heat, solar rays, exercise, and mechanical irritation. The physical urticarias can be distinguished by the precipitating event and other aspects of the clinical presentation. *Dermatographism*, which occurs in 1–4% of the population, is defined by the appearance of a linear wheal with surrounding erythema at the site of a brisk stroke with a firm object ([Fig. 345-3](#)). Dermographism has a prevalence that peaks in the second to third decades. It is not influenced by atopy and has a duration generally of <5 years. *Pressure urticaria*, which often accompanies chronic idiopathic urticaria, presents in response to a sustained stimulus such as a shoulder strap or belt, running (feet), or manual labor (hands). *Cholinergic urticaria* is distinctive in that the pruritic wheals are of small size (1–2 mm) and are surrounded by a large area of erythema; attacks are precipitated by fever, a hot bath or shower, or exercise and are presumptively attributed to a rise in core body temperature. *Exercise-induced anaphylaxis* can be precipitated by exertion alone or can be dependent on prior food ingestion. There is an association with the presence of IgE specific for α -5 gliadin, a component of wheat. The clinical presentation can be limited to flushing, erythema, and pruritic urticaria but may progress to angioedema of the face, oropharynx, larynx, or intestine or to vascular collapse; it is distinguished from cholinergic urticaria by presenting with wheals of conventional size and by not occurring with fever or a hot bath. *Cold urticaria* is local at body areas exposed to low ambient temperature or cold objects but can progress to vascular collapse with immersion in cold water (swimming). *Solar urticaria* is subdivided into six groups by the response to specific portions of the light spectrum. *Vibratory angioedema* may occur after years of occupational exposure or can be idiopathic; it may be accompanied by cholinergic urticaria. Other rare forms of physical allergy, always defined by stimulus-specific elicitation, include *local heat urticaria*, *aquagenic urticaria* from contact with water of any temperature (sometimes associated with polycythemia vera), and *contact urticaria* from direct interaction with some chemical substance (such as latex).

FIGURE 345-3

Dermatographic urticarial lesion induced by stroking the forearm lightly with the edge of a tongue blade. The photograph, taken after 10 min, demonstrates a prominent wheal-and-flare reaction in the shape of a *hashtag*. (Photograph provided by Katherine N. Cahill, MD, Harvard Medical School.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Isolated Angioedema

Angioedema without urticaria can be idiopathic or due to the generation of bradykinin in the setting of C1 inhibitor (C1INH) deficiency that may be inborn as an autosomal dominant characteristic or may be acquired through the appearance of an autoantibody in the setting of malignancy. The angiotensin-converting enzyme (ACE) inhibitors can provoke a similar clinical presentation in 0.2–0.7% of exposed patients due to delayed degradation of bradykinin. Black race, organ transplant, female gender, smoking, and increasing age are known risk factors for ACE-inhibitor related angioedema.

CLINICAL PRESENTATION AND PATHOPHYSIOLOGY

Urticarial eruptions are distinctly pruritic, may involve any area of the body from the scalp to the soles of the feet, and appear in crops of 12- to 36-h duration, with old lesions fading as new ones appear. Most of the physical urticarias (cold, cholinergic, dermatographism) are an exception, with individual lesions lasting <2 h. Neither urticaria nor angioedema lesions are symmetric or dependent in distribution. The most common sites for angioedema are often periorbital and perioral. Angioedema of the upper respiratory tract may be life-threatening due to transient laryngeal obstruction, whereas gastrointestinal involvement may present with abdominal colic, with or without nausea and vomiting, and can result in unnecessary surgical intervention. No residual scarring occurs with either urticaria or angioedema unless there is an underlying vasculitic process.

The pathology is characterized by edema of the superficial dermis in urticaria and of the subcutaneous tissue and deep dermis in angioedema. Collagen bundles in affected areas are widely separated, and the venules are sometimes dilated. Any perivenular infiltrate consists of lymphocytes, monocytes, eosinophils, and neutrophils that are present in varying combination and numbers.

The best evidence for IgE- and mast cell-involvement in urticaria and angioedema is *cold urticaria*. Cryoglobulins or cold agglutinins are present in up to 5% of these patients. Immersion of an extremity in an ice bath precipitates angioedema of the distal portion with urticaria at the air interface within minutes of the challenge. Histologic studies reveal marked mast cell degranulation with associated edema of the dermis and subcutaneous tissues. Elevated levels of histamine have been found in the plasma of venous effluent and in the fluid of suction blisters at experimentally induced lesional sites in patients with cold urticaria, dermatographism, pressure urticaria, vibratory angioedema, light urticaria, and heat urticaria. By ultrastructural analysis, the pattern of mast cell degranulation in cold urticaria resembles an IgE-mediated response with solubilization of granule contents, fusion of the perigranular and cell membranes, and discharge of granule contents, whereas in a dermatographic lesion, there is additional superimposed zonal (piecemeal) degranulation. Elevations of plasma histamine levels with biopsy-proven mast cell degranulation have also been demonstrated with generalized attacks of *cholinergic urticaria*.

Up to 45% of patients with chronic urticaria have an autoimmune cause for their disease including autoantibodies to IgE or to the α chain of Fc ϵ RI. In some patients, autologous serum injected into their own skin can induce a wheal-and-flare reaction involving mast cell activation. The presence of these antibodies can also be recognized by their capacity to release histamine or induce activation markers such as CD63 or CD203 on basophils. An association with antibodies to microsomal peroxidase and/or thyroglobulin has been observed with both clinically significant Hashimoto's thyroiditis as well as a euthyroid state. In vitro studies reveal that these autoantibodies can mediate basophil degranulation with enhancement by serum as a source of the anaphylatoxic fragment, C5a.

The urticaria and angioedema associated with classic serum sickness or with hypocomplementemic cutaneous necrotizing angitis (urticarial vasculitis) are believed to be immune-complex diseases. Reactions to mast cell granule-releasing agents (opioids, contrast media) and to non-steroidal anti-inflammatory drugs are most often limited to urticaria and/or angioedema, but may be systemic.

Hereditary angioedema (HAE) is a fully penetrant, autosomal dominant disease due to a mutation in the *SERPINE1* gene leading to a deficiency of C1INH (type 1) in about 85% of patients or to a dysfunctional protein (type 2) in the remainder affecting 1:30,000–80,000 in the general population. A third less common type of HAE has been described in which C1INH function is normal, and the causal lesion is a mutant form of factor XII, which leads to generation of excessive bradykinin. C1INH deficiency can also develop in a sporadic acquired form as a result of excessive consumption of C1INH due either to formation of immune complexes or to the generation of an autoantibody directed to C1INH in the setting of lymphoproliferative disease. C1INH blocks the catalytic function of activated factor XII (Hageman factor) and of kallikrein, as well as the C1r/C1s components of C1, with the common result of degrading bradykinin. During clinical attacks of angioedema, C1INH function or levels fall, patients develop elevated plasma levels of bradykinin leading to angioedema and excessive activation of C1 results in a decline in C4 and C2 levels.

The use of ACE inhibitors results in impaired bradykinin degradation and explains the angioedema that occurs idiosyncratically in ACE inhibitor-exposed patients with a normal C1INH. Bradykinin-mediated angioedema, whether caused by ACE inhibitors or by C1INH deficiency, is noteworthy for the conspicuous absence of concomitant urticaria or pruritus, the frequent involvement of the gastrointestinal tract, and the duration of symptoms >24 h.

DIAGNOSIS

The classification of urticarial and angioedematous states as presented in [Table 345-1](#) in terms of duration can facilitate identification of possible mechanisms. History alone of self-limited urticarial and/or angioedema episodes can be sufficient to make a diagnosis in the setting of acute disease triggered by drug,

environmental or food allergen with history-directed confirmatory skin testing or assay for serum allergen-specific IgE. Direct reproduction of the lesion in physical urticarias is particularly valuable because it so often establishes the cause of the lesion. Even with chronic urticaria/angioedema, initial diagnostic testing should be limited and expanded testing guided by history. Complete blood count with assessment for eosinophilia, erythrocyte sedimentation rate and thyroid stimulation hormone level are recommended by consensus guidelines even though the vast majority of chronic urticaria is associated with no laboratory abnormality. Urticarial lesions that last longer than 36 h result in scarring and are reported as painful and not pruritic warrant biopsy to evaluate for cellular infiltration, nuclear debris, and fibrinoid necrosis of the venules consistent with urticarial vasculitis. Chronic angioedema without urticaria warrants assessment of complement levels. Concomitant flushing and hyperpigmented papules that urticate with stroking in the absence of angioedema raise the question of mastocytosis. An appropriate travel history should trigger an evaluation for parasites.

The diagnosis of HAE is suggested not only by family history but also by the lack of pruritus and of urticarial lesions, the prominence of recurrent gastrointestinal attacks of colic, and episodes of laryngeal edema. Laboratory diagnosis depends on demonstrating a deficiency of C1INH antigen (type 1) or a nonfunctional protein (type 2) by a catalytic inhibition assay. While levels of C1 are normal, its substrates, C4 and C2, are chronically depleted and fall further during attacks due to the activation of additional C1. Patients with the acquired forms of C1INH deficiency have the same clinical manifestations but differ in the lack of a familial element. Furthermore, their sera exhibit a reduction of C1 function and C1q protein as well as C1INH, C4, and C2. Inborn C1INH deficiency and ACE inhibitor–elicited angioedema are associated with elevated levels of bradykinin. Lastly, type 3 HAE is associated with normal levels of complement proteins and a factor XII gene mutation.

TREATMENT TREATMENT

Urticaria and Angioedema

For most forms of urticaria, H₁ antihistamines such as [chlorpheniramine](#) or [diphenhydramine](#) effectively attenuate both urtication and pruritus, but because of their side effects and short half-life, long-acting, non-sedating agents such as loratadine, desloratadine, and [fexofenadine](#), or low-sedating agents such as [cetirizine](#) or [levocetirizine](#) generally are used first and increased to four times daily (QID) dosing. The addition of an H₂ antagonist such as cimetidine, [ranitidine](#), or [famotidine](#) in conventional dosages may add benefit when H₁ antihistamines are inadequate. A CysLT₁ receptor antagonist such as montelukast, 10 mg daily, or zafirlukast, 20 mg twice a day, can be an important add-on therapy. For chronic urticaria which has failed to respond to a combination of long-acting H₁ antihistamines QID and a CysLT₁ receptor antagonist or cold urticaria, monoclonal anti-IgE antibodies such as [omalizumab](#) are now the next line of therapy. Older agents with antihistamine properties such as [doxepin](#), cyproheptadine, and hydroxyzine have proven effective when H₁ antihistamines fail but are less effective than [omalizumab](#) and are sedating.

Topical glucocorticoids are of no value, and systemic glucocorticoids are generally avoided in idiopathic, allergen-induced, or physical urticarias due to their long-term toxicity. Systemic glucocorticoids are useful in the management of patients with pressure urticaria, vasculitic urticaria (especially with eosinophil prominence), idiopathic angioedema with or without urticaria, or chronic urticaria that responds poorly to conventional treatment and should be considered in any patient with debilitating disease. With persistent vasculitic urticaria, hydroxychloroquine, [dapsone](#), or colchicine may be added to the regimen after hydroxyzine and before or along with systemic glucocorticoids. [Cyclosporine](#) is efficacious for patients with chronic idiopathic urticaria that is severe and poorly responsive to other modalities and/or where glucocorticoids are a requirement.

Infusion of isolated or recombinant C1INH protein is approved for prophylaxis of and acute HAE attacks while administration of a bradykinin 2 receptor antagonist (Icatibant) or a kallikrein inhibitor ([Ecallantide](#)) may be used for treatment of an acute attack of HAE. Older, less expensive preventative options include attenuated androgens, which stimulate production by the normal gene of an amount of functional C1INH sufficient to control the spontaneous activation of C1. The antifibrinolytic agent ε-aminocaproic acid may be used for preoperative prophylaxis, but is contraindicated in patients with thrombotic tendencies or ischemia due to arterial atherosclerosis. Fresh frozen plasma infusion can be used for acute attacks in a setting which lacks access to the newer agents. Bradykinin 2 receptor antagonist and C1INH protein are being studied for ACE inhibitor-induced angioedema. Treatment of the underlying hematologic malignancy is indicated for acquired C1INH deficiency.

ALLERGIC RHINITIS

DEFINITION

Rhinitis is characterized by sneezing; rhinorrhea; obstruction of the nasal passages; conjunctival, nasal, and pharyngeal itching; and lacrimation and can be classified as allergic or non-allergic. A clinical history of rhinitis symptoms occurring in a temporal relationship to allergen exposure and documentation of sensitization to an environmental allergen are required for a diagnosis of allergic rhinitis. Although commonly seasonal due to elicitation by airborne pollens, it can be perennial in an environment of chronic exposure to house dust mites, animal danders, or insect (cockroach) products. The overall prevalence in North America has increased in the past 20 years and is 10–30%, with the peak prevalence of >30% occurring in the fifth decade.

PREDISPOSING FACTORS AND ETIOLOGY

Allergic rhinitis generally occurs in atopic individuals, often in association with atopic dermatitis, food allergy, urticaria, and/or asthma ([Chap. 281](#)). Up to 50% of patients with allergic rhinitis manifest asthma, whereas 70–80% of individuals with asthma and 80% of individuals with chronic bilateral sinusitis experience allergic rhinitis. Female sex, particulate air pollution exposure, and maternal tobacco smoking increase the risk of developing allergic rhinitis over the life span.

Trees, grasses, and weeds that depend on wind rather than insects for pollination produce sufficient quantities of pollen suitable for wide distribution by air currents to elicit seasonal allergic rhinitis. The dates of pollination of these species historically varied little from year to year in a particular locale, but may be quite different in another climate. In the temperate areas of North America, trees typically pollinate from March through May, grasses in June and early July, and ragweed from mid-August to early October. Molds, which are widespread in nature because they occur in soil or decaying organic matter, propagate spores in a pattern that depends on

climatic conditions. Climate change is impacting these patterns with early tree pollination and prolonged ragweed season with the delay of the first frost. Perennial allergic rhinitis occurs in response to allergens that are present throughout the year, including animal dander, cockroach-derived proteins, mold spores, or dust mites such as *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*. Dust mites are scavengers of human skin and excrete [cysteine](#) protease allergens in their feces. In up to 40% of patients with perennial rhinitis, no clear-cut allergen can be demonstrated as causative.

PATHOPHYSIOLOGY AND MANIFESTATIONS

Episodic rhinorrhea, sneezing, obstruction of the nasal passages with lacrimation, and pruritus of the conjunctiva, nasal mucosa, and oropharynx are the hallmarks of allergic rhinitis. The nasal mucosa is pale and boggy, the conjunctiva congested and edematous, and the pharynx generally unremarkable. Swelling of the turbinates and mucous membranes with obstruction of the sinus ostia and eustachian tubes precipitates secondary infections of the sinuses and middle ear, respectively. A growing number of patients with seasonal allergic rhinitis demonstrate pollen-associated food allergen syndrome characterized by oropharyngeal pruritus and/or mild swelling following the ingestion of raw plant-based foods which contain cross-reacting pollen-related allergens.

Nasal polyps, representing mucosal protrusions containing edema fluid with variable numbers of eosinophils and degranulated mast cells, can increase obstructive symptoms with anosmia as a defining feature and can concurrently arise within the nasopharynx or sinuses. Atopy is not a risk factor for nasal polyps, which instead may occur in the setting of cystic fibrosis, aspirin-exacerbated respiratory disease characterized by the triad of asthma, rhinosinusitis, and respiratory reactions to all cyclooxygenase-1 inhibitors, and in patients with chronic staphylococcal colonization, which produces superantigens leading to an intense T_H2 inflammatory response.

The nose presents a large mucosal surface area through the folds of the turbinates and serves to adjust the temperature and moisture content of inhaled air and to filter out particulate materials $>10\ \mu\text{m}$ in size by impingement in a mucous blanket; ciliary action moves the entrapped particles toward the pharynx. Entrapment of pollen and digestion of the outer coat by mucosal enzymes such as lysozymes release protein allergens. The initial interaction occurs between the allergen and intraepithelial mast cells and then proceeds to involve deeper perivascular mast cells, both of which are sensitized with specific IgE. During the symptomatic season when the mucosae are already swollen and hyperemic, there is enhanced adverse reactivity to the seasonal pollen as well as irritants such as tobacco smoke and fragrances. Biopsy specimens of nasal mucosa during seasonal rhinitis show submucosal edema with infiltration by eosinophils, along with some basophils and neutrophils.

The mucosal surface fluid contains IgA that is present because of its secretory piece and also IgE, which apparently arrives by diffusion from plasma cells in proximity to mucosal surfaces. IgE fixes to mucosal and submucosal mast cells, and the intensity of the clinical response to inhaled allergens is quantitatively related to the naturally occurring pollen dose. In sensitive individuals, the introduction of allergen into the nose is associated with sneezing, nasal obstruction, and discharge, and the fluid contains histamine, PGD_2 , and leukotrienes. Thus the mast cells of the nasal mucosa and submucosa generate and release mediators through IgE-dependent reactions that are capable of producing tissue edema and eosinophilic infiltration.

DIAGNOSIS

The diagnosis of seasonal allergic rhinitis depends largely on an accurate history of occurrence coincident with the pollination of the offending weeds, grasses, or trees. The continuous character of perennial allergic rhinitis due to contamination of the home or place of work makes historic analysis difficult, but there may be variability in symptoms that can be related to exposure to animal dander, dust mite and/or cockroach allergens, fungal spores, or work-related allergens such as latex. Patients with perennial rhinitis commonly develop the problem in adult life, and manifest nasal congestion and a postnasal discharge, often associated with thickening of the sinus membranes demonstrated by radiography. Perennial nonallergic rhinitis with eosinophilia syndrome (NARES) occurs in the middle decades of life and is characterized by nasal obstruction, anosmia, chronic sinusitis, and prominent eosinophilic nasal discharge in the absence of allergen sensitization. The term *vasomotor rhinitis* or *perennial nonallergic rhinitis* designates a condition of enhanced reactivity of the nasopharynx in which a symptom complex resembling perennial allergic rhinitis occurs with nonspecific stimuli, including chemical odors, temperature and humidity variations, and position changes but occurs without tissue eosinophilia or an allergic etiology. Other entities to be excluded are structural abnormalities of the nasopharynx; exposure to irritants; gustatory rhinitis associated with cholinergic activation that occurs while eating or ingesting alcohol; hypothyroidism; upper respiratory tract infection; pregnancy with prominent nasal mucosal edema; prolonged topical use of α -adrenergic agents in the form of nasal sprays (rhinitis medicamentosa); and the use of certain systemic agents such as β -adrenergic antagonists, ACE inhibitors, direct vasodilators (hydralazine), $\alpha1$ -adrenergic receptor antagonists, [estrogens](#), progesterone, NSAIDs, gabapentin, phosphodiesterase-5 inhibitors, and psychotropics (Risperidone, chlorpromazine, amitriptyline).

The nasal secretions of allergic patients are rich in eosinophils, and a modest peripheral eosinophilia can be observed. Local or systemic neutrophilia implies infection. Total serum IgE is frequently elevated, but the demonstration of immunologic specificity for IgE is critical to an etiologic diagnosis. A skin test by the intracutaneous route (puncture or prick) with the allergens of interest provides a rapid and reliable approach to identifying allergen-specific IgE that has sensitized cutaneous mast cells. A positive intracutaneous skin test with 1:10–1:20 weight/volume of extract has a high predictive value for the presence of allergy. An intradermal test with a 1:500–1:1000 dilution of 0.05 mL may follow if indicated by history when the intracutaneous test is negative, but while more sensitive, it is less reliable due to the reactivity of some asymptomatic individuals at the test dose.

Newer methodology for detecting total IgE, including the development of enzyme-linked immunosorbent assays (ELISA) employing anti-IgE bound to either a solid-phase or a liquid-phase particle, provides rapid and cost-effective determinations. Measurements of specific anti-IgE in serum are obtained by its binding to an allergen and quantitation by subsequent uptake of labeled anti-IgE. As compared to the skin test, the assay of specific IgE in serum is less sensitive but has high specificity.

TREATMENT

TREATMENT

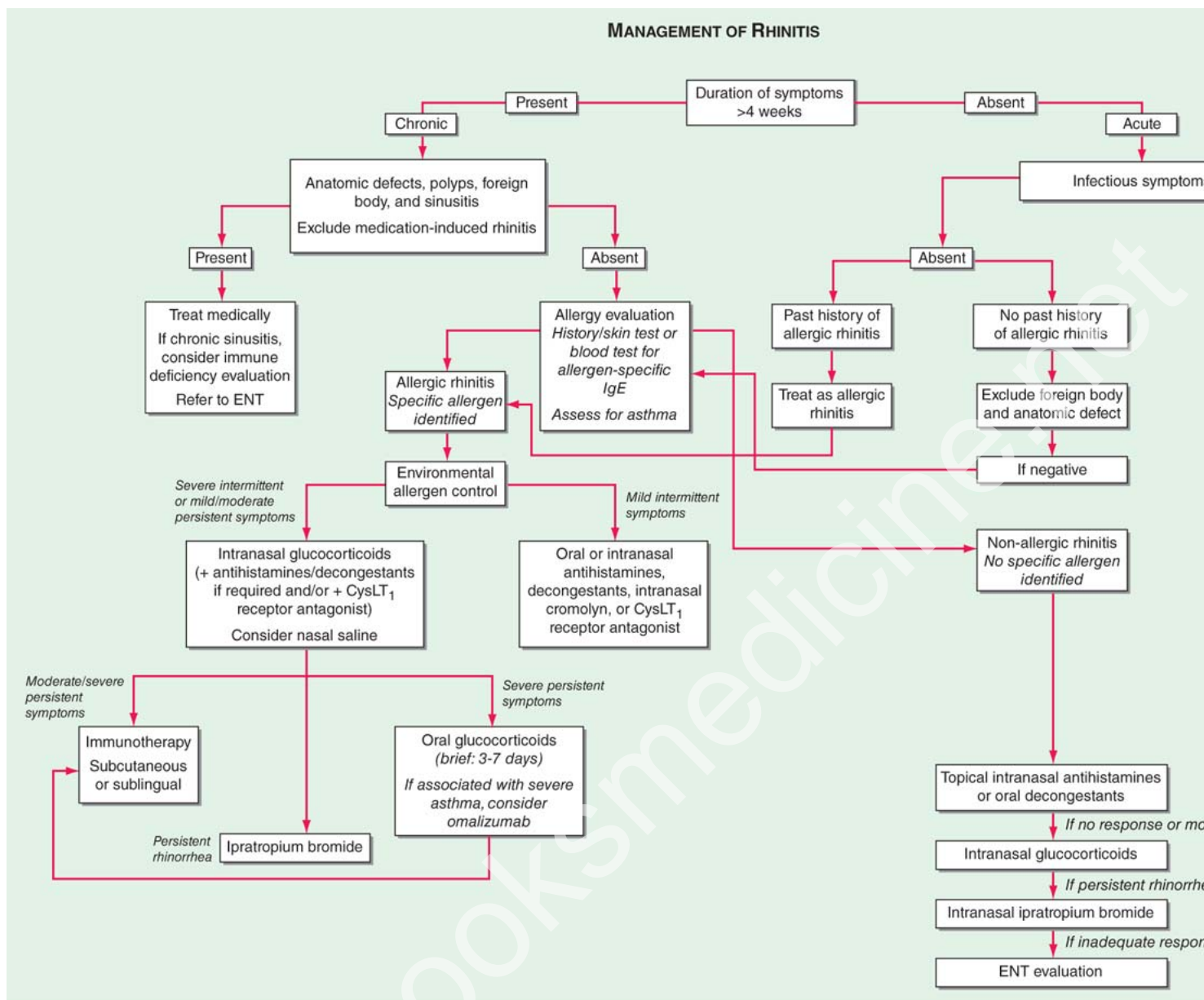
Allergic Rhinitis

Although allergen avoidance is the most cost-effective means of managing allergic rhinitis, only in the case of animal dander and possibly dust mites is it really feasible. Treatment with pharmacologic agents represents the standard approach to seasonal or perennial allergic rhinitis. Oral long-acting H₁ antihistamines are effective for nasopharyngeal itching, sneezing, and watery rhinorrhea and for such ocular manifestations as itching, tearing, and erythema, but they are less efficacious for the nasal congestion. The older antihistamines are sedating, and they induce psychomotor impairment, including reduced eye-hand coordination and impaired automobile driving skills. Their anticholinergic (muscarinic) effects include visual disturbance, urinary retention, and constipation. Because the newer H₁ antihistamines such as **fexofenadine**, loratadine, desloratadine, **cetirizine**, **levocetirizine**, olopatadine, bilastine, and **azelastine** are less lipophilic and more H₁ selective, their ability to cross the blood-brain barrier is reduced, and thus their sedating and anticholinergic side effects are minimized. These newer antihistamines do not differ appreciably in efficacy for relief of rhinitis and/or sneezing. Intranasal high-potency glucocorticoids are the most potent drugs available for the relief of established rhinitis, seasonal or perennial, and are effective in relieving nasal congestion as well as ocular symptoms. They provide efficacy with substantially reduced side effects as compared with this same class of agent administered orally. Their most frequent side effect is local irritation, with *Candida* overgrowth being a rare occurrence. The currently available intranasal glucocorticoids—beclomethasone, **flunisolide**, **triamcinolone**, **budesonide**, **fluticasone** propionate, **fluticasone** furoate, ciclesonide, and **mometasone** furoate—are equally effective for nasal symptom relief, including nasal congestion; these agents all achieve up to 70% overall symptom relief with some variation in the time period for onset of benefit. **Azelastine** nasal spray may benefit individuals with nonallergic vasomotor rhinitis as well as additive benefit to intranasal steroids in allergic rhinitis, but it has an adverse effect of dysgeusia (taste perversion) in some patients. Alternative nasal decongestants include α -adrenergic agents such as **phenylephrine** or **oxymetazoline**; however, the duration of their efficacy is limited because of rebound rhinitis (i.e., 7- to 14-day use can lead to rhinitis medicamentosa) and such systemic responses as hypertension. Oral α -adrenergic agonist decongestants containing pseudoephedrine are standard for the management of nasal congestion, generally in combination with an antihistamine. While oral antihistamines typically reduce nasal and ocular symptoms by about one-third, pseudoephedrine must be added to achieve a similar reduction in nasal congestion. These pseudoephedrine combination products can cause insomnia and are precluded from use in patients with narrow angle glaucoma, urinary retention, severe hypertension, marked coronary artery disease, or a first-trimester pregnancy. The CysLT₁ blocker montelukast is approved for treatment of both seasonal and perennial rhinitis, and it reduces both nasal and ocular symptoms by about 20%. **Cromolyn** sodium nasal spray inhibits mast cell degranulation, and can be used prophylactically on a continuous basis during the season. Topical **ipratropium** is an anticholinergic agent effective in reducing rhinorrhea, including that of patients with perennial non-allergic symptoms, and it can be additionally efficacious when combined with intranasal glucocorticoids. For concomitant allergic conjunctivitis, topical treatment with **cromolyn** sodium is effective in treating mild allergic symptoms and topical antihistamines such as olopatadine, **azelastine**, **ketotifen**, or **epinastine** administered to the eye provide rapid relief of itching and redness and are more effective than oral antihistamines.

Immunotherapy Immunotherapy consists of repeated exposure to gradually increasing concentrations of the allergen(s) considered to be specifically responsible for the symptom complex. Two forms of immunotherapy, subcutaneous (SCIT) and sublingual (SLIT), are currently available. Controlled studies of ragweed, grass, dust mite, and cat dander allergens administered via SCIT for treatment of allergic rhinitis have demonstrated improved symptom control over medications alone with the advantage of providing a durable benefit. The duration of SCIT is 3–5 years, with discontinuation being based on minimal symptoms over two consecutive seasons of exposure to the allergen. Clinical benefit appears related to the administration of a high dose of relevant allergen, advancing from weekly to monthly intervals. Patients should remain at the treatment site for at least 30 min after allergen administration so that any systemic reactions including anaphylaxis can be managed. Two to three percent of SCIT patients experience a systemic reaction over a 12-month period. Local reactions with erythema and induration are not uncommon and may persist for 1–3 days. SLIT is prepared as a tablet to be dissolved under the tongue at home after the first dose. The efficacy of SLIT is comparable to SCIT but only for the three allergens formulations available, dust mite, timothy/northern grasses and ragweed. Systemic reactions are less frequent with SLIT but transient oral pruritus is common. Immunotherapy is contraindicated in patients with significant cardiovascular disease or unstable asthma and should be conducted with particular caution in any patient requiring β -adrenergic blocking therapy because of the difficulty in managing an anaphylactic complication. The response to immunotherapy is associated with a complex of cellular and humoral effects that includes a modulation in T cell cytokine production and allergen-specific IgG₄ expansion. Immunotherapy should be reserved for clearly documented seasonal or perennial rhinitis that is clinically related to defined allergen exposure with confirmation by the presence of allergen-specific IgE through skin or in vitro specific IgE testing. Systemic treatment with a monoclonal antibody to IgE (**omalizumab**) that blocks mast cell and basophil sensitization has efficacy for allergic rhinitis and can be used with immunotherapy to enhance safety and efficacy. However, current approval is only for treatment of patients with persistent allergic asthma not controlled by inhaled glucocorticoid therapy. A sequence for the management of allergic or perennial rhinitis based on an allergen-specific diagnosis and stepwise management as required for symptom control would include the following: (1) identification of the offending allergen(s) by history with confirmation of the presence of allergen-specific IgE by skin test and/or serum assay; (2) avoidance of the offending allergen; and (3) medical management in a stepwise fashion (**Fig. 345-4**). Mild intermittent symptoms of allergic rhinitis are treated with oral antihistamines, oral CysLT₁ receptor antagonists, intranasal antihistamines, or intranasal **cromolyn** prophylaxis. Moderate to more severe allergic rhinitis is managed with intranasal glucocorticoids plus oral antihistamines, oral CysLT₁ receptor antagonists, or antihistamine-decongestant combinations. Persistent or seasonal allergic rhinitis, rhinoconjunctivitis, or asthma which remains uncontrolled with maximal medical therapy merit consideration of allergen-specific immunotherapy.

FIGURE 345-4

Algorithm for the diagnosis and management of rhinitis. Persistent defined as >4 days/week for >4 weeks. Moderate/severe defined as abnormal sleep, impaired daily activities (school, work, sport, leisure) and/or troublesome symptoms. CysLT, cysteinyl leukotriene; ENT, ear, nose, and throat; IgE, immunoglobulin E.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

FURTHER READING

Bernstein DI et al: Allergic rhinitis: Mechanisms and treatment. *Immunol Allergy Clin North Am* 36:261, 2016. [PubMed: 27083101]

Cho SH et al: Chronic rhinosinusitis without nasal polyps. *J Allergy Clin Immunol Pract* 4:575, 2016. [PubMed: 27393771]

Cicardi M et al: Classification, diagnosis, and approach to treatment for angioedema: Consensus report from the Hereditary Angioedema International Working Group. *Allergy* 69:602, 2014. [PubMed: 24673465]

Corren J et al: Allergic and nonallergic rhinitis, in *Middleton's Allergy: Principles and Practice*, 8th ed. NF Adkinson et al (eds). Philadelphia, Saunders, 2014, pp 664–685.

Jutel M et al: International consensus on allergen immunotherapy II: Mechanisms, standardization, and pharmacoeconomics. *J Allergy Clin Immunol* 137:358, 2016. [PubMed: 26853128]

Maurer M et al: Omalizumab for the treatment of chronic idiopathic or spontaneous urticaria. *N Engl J Med* 368:924, 2013. [PubMed: 23432142]

Saini SS: Urticaria and angioedema, in *Middleton's Allergy: Principles and Practice*, 8th ed. NF Adkinson et al (eds). Philadelphia, Saunders, 2014, pp 575–587.

Stevens WW et al: Chronic rhinosinusitis with nasal polyps. J Allergy Clin Immunol Pract 4:565, 2016. [[PubMed: 27393770](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 346: Anaphylaxis

David Hong; Joshua A. Boyce

DEFINITION

Anaphylaxis is a potentially life-threatening systemic allergic reaction involving one or more organ systems that typically occurs within seconds to minutes of exposure to the anaphylactic trigger, most often a drug, food, or hymenoptera sting. Other triggers of anaphylaxis include radiocontrast administration or latex exposure. The term “anaphylaxis” was first described in 1902 by Charles Richet and Paul Portier who attempted to immunize dogs against sea anemone toxin in the same way Pasteur was able to vaccinate individuals against the smallpox virus. To their surprise, repeated administration of small, sub-lethal doses of sea anemone toxin reliably induced acute-onset death when re-administered 2–3 weeks after initial “vaccination” to the toxin. The phenomenon was termed ana (anti) - phylaxis (“protection or guarding”) because vaccination with anemone toxin resulted in the opposite intended immune effect. Charles Richet was awarded the Nobel Prize in Physiology or Medicine in 1913 for this work which led to further insights into hypersensitivity and mast cell biology.

Clinical Manifestations

While 80–90% of anaphylactic episodes are uniphasic, about 10–20% of cases are biphasic in which anaphylactic symptoms return about an hour or longer after resolution of initial symptoms. Anaphylactic reactions are particularly dangerous when hypotension or hypoxia occurs, leading potentially to cardiovascular collapse or respiratory failure, respectively. There may be upper or lower airway obstruction or both. Laryngeal edema may be experienced as a “lump” in the throat, hoarseness, or stridor, whereas bronchial obstruction is associated with a feeling of tightness in the chest and/or audible wheezing. Patients with underlying asthma are predisposed to severe involvement of the lower airways and increased mortality associated with anaphylaxis. In fatal cases with clinical bronchial obstruction, the lungs show marked hyperinflation on gross and microscopic examination. The microscopic findings in the bronchi, however, are limited to luminal secretions, peribronchial congestion, submucosal edema, and eosinophilic infiltration, and the acute emphysema is attributed to intractable bronchospasm that subsides with death. Angioedema resulting in death by mechanical obstruction occurs in the epiglottis and larynx; however, the process also is evident in the hypopharynx and to some extent in the trachea. On microscopic examination, there is wide separation of the collagen fibers and the glandular elements; vascular congestion and eosinophilic infiltration also are present. Patients dying of vascular collapse without antecedent hypoxia from respiratory insufficiency have visceral congestion with a presumptive loss of intravascular fluid volume. The associated electrocardiographic abnormalities, with or without infarction, in some patients may reflect a primary

cardiac event mediated by mast cells (which are prominent near the coronary vessels) or may be secondary to a critical reduction in blood volume.

Gastrointestinal manifestations represent another severe presentation of anaphylaxis, and include nausea, vomiting, crampy abdominal pain, and/or fecal incontinence. Angioedema of the bowel wall may also cause sufficient intravascular volume depletion to precipitate cardiovascular collapse.

Cutaneous manifestations are among the most common presentations of anaphylaxis (>90% of cases). Symptoms include urticarial eruptions, flushing with diffuse erythema, and/or a feeling of generalized warmth. Urticarial eruptions are intensely pruritic and may be localized or disseminated. They may coalesce to form giant hives but seldom persist beyond 48 h.

PREDISPOSING FACTORS AND ETIOLOGY

Because the most dangerous manifestations of anaphylaxis involve the cardiovascular and/or respiratory systems, preexisting asthma and underlying cardiovascular disease could lead to more rapid decompensation from anaphylaxis. Atopy is not generally thought to be a risk factor for anaphylaxis from drug reactions or hymenoptera stings, but is associated with radiocontrast sensitivity, exercise-induced anaphylaxis, idiopathic anaphylaxis, and allergy to foods or latex. Severe hymenoptera-induced anaphylaxis (generally with prominent hypotension) can be a presenting feature of underlying systemic mastocytosis. Hymenoptera allergy is also more likely in patients whose occupations (i.e., beekeepers, trash haulers, and landscape workers) place them in regular proximity to stinging insects. Most commonly, allergen-induced cross-linking of IgE-bound FcεRI receptors on mast cells and basophils initiates the signal transduction events leading to hypersensitivity syndromes including anaphylaxis. The generation of allergen-specific IgE is the end result of sensitization via the adaptive immune system. The mechanisms underlying sensitization are beyond the scope of this topic; however, environmental factors, innate immune responses, and cytokines are among the many variables leading to antigen-specific IgE production by B cells and plasma cells. IgE-mediated drug allergies are most common with antibiotics and certain chemotherapy drugs, though theoretically, they can occur with almost any medication. As is the case with environmental allergies, repeated exposure to the allergy-causing antigen is an important risk factor to keep in mind when evaluating patients with anaphylaxis. In the case of allergy to carboplatin, the incidence of hypersensitivity is 27% in patients who have had ≥7 lifetime infusions and as high as 46% in patients who have had ≥15 lifetime infusions. Similarly, patients with cystic fibrosis have a relatively high incidence of allergic reactions to IV antibiotics that they receive periodically to treat exacerbations of bronchiectasis. Drugs can also function as haptens that form immunogenic conjugates with host proteins. The conjugating hapten may be the parent compound, a nonenzymatically derived storage product, or a metabolite formed in the host. Recombinant biologics can also induce the formation of IgE against the proteins or against glycosylated structures that serve as immunogens. More recently, outbreaks of anaphylaxis to the EGFR antibody, cetuximab, were reported in association with elevated titers of serum IgE to alpha-1,3-galactose (alpha-gal), an oligosaccharide found in non-primate mammals. Cetuximab is derived from a mouse cell line expressing a transferase that tags the Fab' portion of the cetuximab heavy chain with alpha-gal. Interestingly, patients

with a history of multiple bites from *Amblyomma americanum* ticks commonly found in the Carolinas, Arkansas, and Tennessee are more likely to have anti-alpha-gal IgE as compared to control patients living outside those states. Such individuals who become sensitized to alpha-gal can develop episodes of delayed anaphylaxis to beef, lamb, and pork.

PATHOPHYSIOLOGY

Many of the important early mediators of anaphylaxis are derived from mast cells, basophils, and eosinophils. Mast cells and basophils contain preformed granules comprised of histamine, proteases (tryptase, chymase), proteoglycans (heparin, chondroitin sulfate), and TNF- α , which are rapidly released into surrounding tissue upon cell activation, a process known as degranulation. Mast cells, basophils, and eosinophils are also sources of arachidonic acid-derived products which include cysteinyl leukotrienes, prostaglandins, and platelet activating factor (PAF). Histamine release results in flushing, urticaria, pruritus, and, in high concentrations, hypotension and tachycardia. Cysteinyl leukotrienes and prostaglandin D₂ cause bronchoconstriction and increased microvascular permeability. Prostaglandin D₂ causes cutaneous flushing, and attracts eosinophils and basophils to the site of mast cell activation. Serum PAF levels correlate with anaphylaxis severity and are inversely proportional to the constitutive level of PAF acetylhydrolase, which is necessary for PAF inactivation. Tryptase and chymase can activate complement and coagulation pathways. Activation of these pathways results in production of the anaphylotoxins, C3a and C5a, and activation of the kallikrein-kinin system which regulates blood pressure and vascular permeability. The actions of these anaphylactic mediators are likely additive or synergistic at the target tissues.

Non-IgE-mediated reactions to certain drugs (which may occur upon the first exposure) can mimic the pathophysiology of IgE-dependent anaphylaxis due to a similar profile of mediators. For example, paclitaxel is a chemotherapy agent derived from yew tree bark and needles that requires polyethoxylated castor oil (Cremophor) to be solubilized into aqueous solution. Cremophor directly activates the complement cascade, resulting in complement-dependent induced histamine release from mast cells and basophils. A version of paclitaxel that is solubilized by being bound to albumin nanoparticles, Abraxane, has a far lower rate of hypersensitivity, especially for patients who have had infusion reactions to Cremophor-solubilized paclitaxel. Reactions to radiocontrast and vancomycin are other examples of non-IgE-mediated hypersensitivity. Opiates and NSAIDs are other drug categories that can have similar adverse reactions.

DIAGNOSIS

The diagnosis of an anaphylactic reaction depends primarily on a history revealing the onset of symptoms and signs within seconds to minutes after the putative trigger is encountered. An exception is delayed anaphylaxis to meats in alpha-gal sensitized patients. Every attempt to identify the specific cause or causes should be made so as to minimize the risk of recurrent anaphylaxis. If a particular drug or food is suspected, skin or serum specific IgE testing is useful to confirm clinical suspicions. If a specific trigger cannot be identified, a workup of underlying atopic diatheses may be useful to identify risk factors that could play a

potential contributory role. In the acute setting, laboratory biomarkers of mast cell degranulation may be useful to document the severity of an anaphylactic episode. The most obvious serum biomarker to assay, histamine, has an extremely short half-life with a measurable time-window that expires <1 h from the onset of anaphylaxis. A more practical and useful biomarker is serum tryptase which peaks 60–90 min after the onset of anaphylaxis and can be measured as long as 5 h after the onset of anaphylaxis. It may be useful to follow-up an elevated tryptase measurement in the acute setting with another measurement when the patient is clinically stable to establish a baseline reference. An elevated baseline tryptase level may warrant further workup for mastocytosis, especially if the presenting reaction occurred in the setting of hymenoptera sting.

TREATMENT

Early recognition of an anaphylactic reaction is mandatory since severe, even fatal, complications, can occur within minutes after symptoms first appear. The treatment of first choice is intramuscular administration of 0.3–0.5 mL of 1:1000 (1 mg/mL) [epinephrine](#), with repeated doses at 5–20 min intervals as needed for a severe reaction. The failure to use [epinephrine](#) within the first 20 min of symptoms is a risk factor for poor clinical outcomes in various studies of anaphylaxis. Another important variable that may affect anaphylaxis survival is body posture, as an upright or sitting posture may lead to the “empty heart syndrome” in which there is insufficient venous return to the heart from sudden onset hypotension secondary to intravascular volume depletion. [Epinephrine](#) can further accelerate empty heart syndrome due to its chronotropic effects. For this reason, it is recommended that patients who suffer from anaphylaxis be placed in the supine position before receiving [epinephrine](#). IV fluids and vasopressor agents may be administered in the acute medical setting if intractable hypotension occurs. [Epinephrine](#) provides both α - and β -adrenergic effects, resulting in vasoconstriction, bronchial smooth-muscle relaxation, and attenuation of enhanced venular permeability. Beta blockers may attenuate this response; therefore, an alternative anti-hypertensive may be considered in patients at high risk of needing emergency [epinephrine](#). Oxygen alone via a nasal catheter or with nebulized albuterol may be helpful; however, either endotracheal intubation or a tracheostomy is mandatory for oxygen delivery if progressive hypoxia develops. Ancillary agents such as antihistamines, glucocorticoids, and bronchodilators are also useful therapeutics to treat urticaria/angioedema and bronchospasm once the patient is hemodynamically stable.

PREVENTION

Avoidance

The simplest, most straightforward approach to the long-term management of a patient with a history of anaphylaxis is strict avoidance of known anaphylactic triggers and education on acute management, that is, instructing the patient on the proper use and indications for use of self-administered [epinephrine](#). Lifelong avoidance is not easy if the trigger is an occupational exposure, hymenoptera sting, a common food (i.e., peanut), or a drug representing the sole or best therapeutic option for the patient. Special management options may exist for these patients.

Patients with large local reactions to hymenoptera stings are unlikely to have anaphylaxis with subsequent stings; however, patients of any age who have had documented anaphylaxis should be formally evaluated and started on venom immunotherapy (VIT) if skin or serologic IgE testing confirms the history. Immunotherapy is a means of “tolerizing” patients to allergen by means of serial subcutaneous administration of escalating doses of extract containing relevant allergen until a target maintenance dose is achieved. As in the case of Richet’s unfortunate dogs, anaphylaxis can sometimes occur during the course of administering immunotherapy extracts, so formulating extracts and administering them is typically done under the care of a specialist familiar with this type of treatment. In the case of hymenoptera allergy, patients receive VIT extracts containing actual hymenoptera venom with a maintenance dose equivalent to 2–5 stings. The recommended duration of treatment is 3–5 years; however, patients who have experienced severe respiratory or cardiovascular anaphylaxis are often on lifelong therapy.

Tolerance Induction

IgE sensitization to foods occurs most frequently in infants and young children, especially those with atopic dermatitis, and is a risk factor for anaphylaxis (although detection of specific IgE through skin or serum testing has relatively poor predictive value). While most allergy to egg, milk, soy, and/or wheat resolves spontaneously during childhood, ~80% of children with peanut allergy remain sensitive for life. A sharp rise in the prevalence of peanut allergy was also observed in the late 1990s–early 2000s especially in countries with Western diets where the average age of peanut introduction was age ≥ 3 years. Curiously, in cultures where peanut was introduced much earlier into children’s diets, the prevalence of peanut allergy remained low. The landmark “Learning Early About Peanut Allergy” (LEAP) study demonstrated that early introduction of peanut protein to the diet of high risk infants (4–11 months of age with atopic dermatitis and/or egg allergy) can prevent the development of most (80% or more) peanut allergy compared with children who did not consume peanuts (avoidance group), even when IgE sensitization (based on positive skin test) had already developed at the time of study entry. While the induction of tolerance at an early age seems to be key to preventing clinical reactivity later in life, it is not yet clear if this principle holds true for other foods commonly associated with hypersensitivity reactions.

Desensitization

For patients who have suffered anaphylaxis from drug allergy and whose treatment regimen requires the administration of the offending drug, desensitization may be a short-term treatment option to prevent reactions. Desensitization elicits a temporary state of tolerance to the drug in sensitized, clinically reactive patients. While it has been a proven technique for penicillin-allergic patients for decades, desensitization has more recently been proven to be effective for certain chemotherapy agents, especially platin-based chemotherapy agents which can induce IgE-mediated sensitization with repeated exposures. The exact mechanisms underlying desensitization are not fully understood; however, temporary tolerance can be achieved through the serial administration of gradually escalating doses of drug, starting from extremely low doses, over the course of hours. So long as the patient continues to receive the drug in question at regular intervals based on drug half-life, a “desensitized” state can also be maintained until the drug is no longer needed. Drug desensitization works best for IgE-mediated reactions; however, it has been performed in cases of non-IgE-mediated anaphylaxis from Cremophor-solubilized paclitaxel as described earlier in this chapter. Other non-IgE-mediated anaphylactic reactions can often be prevented with premedication regimens. A

typical premedication regimen for radiocontrast, for example, will have the patient receive [prednisone](#) 0.5 mg/kg at 13, 6, and 1 h prior to contrast administration. [Diphenhydramine](#) 25 mg is also given 1 h prior to contrast. Flushing reactions from vancomycin can often be alleviated with antihistamine premedication and down titrating the infusion rate.

FURTHER READING

Castells MC et al: Hypersensitivity reactions to chemotherapy: Outcomes and safety of rapid desensitization in 413 cases. *J Allergy Clin Immunol* 122:574, 2008. [[PubMed: 18502492](#)]

Chung CH et al: Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *N Engl J Med* 358:1109, 2008. [[PubMed: 18337601](#)]

Du Toit G et al: LEAP Study Team. Randomized trial of peanut consumption in infants at risk for peanut allergy. *N Engl J Med* 373:803, 2015.

Du Toit G et al: Immune Tolerance Network LEAP-On Study Team. Effect of avoidance on peanut allergy after early peanut consumption. *N Engl J Med* 374:1435, 2016. [[PubMed: 26942922](#)]

Lieberman P et al: Anaphylaxis—A practice parameter update 2015. *Ann Allergy Asthma Immunol* 115:341, 2015. [[PubMed: 26505932](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 347: Mastocytosis

Cem Akin; Joshua A. Boyce

DEFINITION

Mastocytosis is defined by accumulation of clonally expanded mast cells in tissues such as skin, bone marrow, liver, spleen, and gut. The mast cell expansion is generally recognized in skin and/or bone marrow. Mastocytosis occurs at any age and has a slight preponderance in males. Mastocytosis is a rare disorder and its exact prevalence is not known; however, it is estimated to occur in ~1 in 20,000 people. Familial occurrence is rare, and atopy is not increased compared to the general population.

CLASSIFICATION AND PATHOPHYSIOLOGY

A consensus classification for mastocytosis recognizes cutaneous mastocytosis with variants, five systemic forms, and rare mast cell sarcoma ([Table 347-1](#)).

TABLE 347-1

Classification of Mastocytosis

Cutaneous mastocytosis (CM)
Maculopapular cutaneous mastocytosis (MPCM)
Solitary mastocytoma of skin
Diffuse cutaneous mastocytosis
Indolent systemic mastocytosis (ISM)
Smoldering systemic mastocytosis
Systemic mastocytosis with an associated clonal hematologic non–mast cell lineage disease (SM-AHNMD)
Aggressive systemic mastocytosis (ASM)
Mast cell leukemia (MCL)
Mast cell sarcoma (MCS)

Source: Modified from H-P Horny et al: Mastocytosis. In: *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, revised 4th ed. SH Swerdlow et al (eds). Lyon, France, IARC Press, 2017, pp 61–69.

Cutaneous mastocytosis is the most common diagnosis of mastocytosis in children and indicates disease limited to skin with absence of pathologic infiltrates in internal organs. It is usually diagnosed within the first year of life with demonstration of fixed, maculopapular, and hyperpigmented lesions (maculopapular cutaneous mastocytosis [MPCM], formerly known as urticaria pigmentosa), mastocytoma(s) or diffuse cutaneous mastocytosis. Systemic mastocytosis (SM) refers to involvement of a non-cutaneous site (usually bone marrow). There are five distinct variants of SM; the form designated as *indolent systemic mastocytosis* (ISM) accounts for the majority of adult patients. ISM is diagnosed when there is no evidence of an associated hematologic disorder, mast cell leukemia or tissue dysfunction due to mast cell infiltration and is not known to alter life expectancy. Systemic smoldering mastocytosis (formerly considered a subvariant of ISM) is characterized by high mast cell burden as evidenced by a bone marrow infiltration of >30% and a baseline serum tryptase >200 ng/ml (B-findings), but absence of SM-AHNMD or ASM ([Table 347-2](#)). In *systemic mastocytosis associated with clonal hematologic non-mast cell lineage disease* (SM-AHNMD, or SM-AHN for short), the prognosis is determined by the nature of the associated disorder, which can range from dysmyelopoiesis to leukemias usually of myeloid origin. In *aggressive systemic mastocytosis* (ASM), mast cell infiltration/proliferation in multiple organs such as liver, spleen, gut, bone, and bone marrow resulting in 1 or more C findings and a poor prognosis ([Table 347-2](#)). *Mast cell leukemia* (MCL) is the rarest form of SM and is invariably fatal at present; the peripheral blood contains circulating, metachromatically staining, and atypical mast cells. An aleukemic form of MCL is recognized without circulating mast cells when the percentage of high-grade immature mast cells in bone marrow smears exceeds 20% in a nonspicular area. Mast cell sarcoma is a rare solid mast cell tumor with malignant invasive features.

TABLE 347-2

B and C Findings for Diagnosis of SSM and ASM

B-Findings (2 or more in the absence of any C findings are required for a diagnosis of SSM):

1. MC infiltration in bone marrow biopsy of >30% and the basal serum tryptase level >200 ng/mL
2. Hypercellular bone marrow with signs of dysmyelopoiesis but without cytopenias meeting C criteria or WHO criteria for an MDS or MPN
3. Palpable hepatomegaly, palpable splenomegaly, or lymphadenopathy (on CT or ultrasound: >2 cm) without impaired liver function or hypersplenism

C-Findings: (1 or more required for a diagnosis of ASM). C finding should be reasonably attributable to high tissue mast cell infiltration.

1. Cytopenia(s): ANC<1,000/ μ L or Hb<10 g/dL or PLT<100,000/ μ L
2. Hepatomegaly with ascites and impaired liver function
3. Palpable splenomegaly with associated hypersplenism
4. Malabsorption with hypoalbuminemia and weight loss
5. Skeletal lesions: large area(s) of osteolyses with pathologic fractures (presence of osteoporosis alone without osteolytic lesions does not satisfy this criterion)

Abbreviations: ANC, absolute neutrophil count; ASM, aggressive systemic mastocytosis; CT, computed tomography; Hb, hemoglobin; MC, mast cells; MDS, myelodysplastic syndromes; MPN, myeloproliferative disorders; PLT, platelets; SSM, smoldering systemic mastocytosis; WHO, World Health Organization.

A point mutation of A to T at codon 816 of *KIT* that causes an aspartic acid to valine substitution, resulting in a somatic gain-of-function mutation, is found in mast cells and sometimes in multiple other cell lineages in patients with mastocytosis. This substitution, as well as other rare mutations of *KIT*, is characteristic of patients with all forms of SM, but is also present in some children with cutaneous mastocytosis in lesional skin, as might be anticipated because mast cells are of bone marrow lineage. Additional mutations in genes such as *TET2*, *SRSF2*, *ASXL1*, and *RUNX1* known to be associated with other hematologic neoplastic disorders can be detected in patients usually with advanced (non-ISM) forms of SM. The prognosis for patients with cutaneous mastocytosis and for almost all patients with ISM is a normal life expectancy, whereas that for patients with SM-AHNMD is determined by the non-mast cell component. ASM and MCL carry a poorer prognosis, while patients with SSM carry an intermediate prognosis. Progression from ISM to a more advanced form is rare (approximately 3% overall); however, patients should be monitored for emergence of hematologic disease and end organ manifestations of ASM. In infants and children with cutaneous manifestations, namely, maculopapular cutaneous mastocytosis, mastocytoma(s), or bullous lesions, visceral involvement is usually lacking, and spontaneous resolution is common prior to adolescence. Progression from CM to ISM may occur in ~10% of children, especially in those with high mast cell burden

(diffuse cutaneous mastocytosis), hematologic abnormalities and those who present with smaller uniform lesions with diameters measuring <2 cm.

CLINICAL MANIFESTATIONS

The clinical manifestations of SM, distinct from a leukemic complication, are due to the release of bioactive substances acting at both local and distal sites, tissue occupancy by the mast cell mass, and the tissue response to that mass. The pharmacologically induced manifestations are intermittent flushing, tachycardia and vascular collapse, gastric distress, lower abdominal crampy pain, and diarrhea. The increase in local cell burden is evidenced by the lesions of MPCM (urticaria pigmentosa) at skin sites and internal organ biopsies such as bone marrow and gastrointestinal tract and may be a direct local cause of bone pain and/or malabsorption. Mast cell-mediated fibrotic changes may occur in liver, spleen, and bone marrow but not in gastrointestinal tissue or skin. Immunofluorescent analysis of bone marrow and skin lesions in ISM and of spleen, lymph node, and skin in ASM has revealed only one mast cell phenotype, namely, scroll-poor cells expressing tryptase, chymase, and CPA.

The cutaneous lesions of MPCM (formerly known as urticaria pigmentosa) are reddish-brown macules, papules, or plaques that respond to trauma with urtication and erythema (Darier's sign). Children with CM may present with MPCM, mastocytomas, or diffuse cutaneous mastocytosis (DCM). Mastocytomas are generally solitary elevated lesions that are yellow, brown, or red in color. Their size may vary from a few millimeters to several centimeters. Rubbing or irritation of the mastocytoma lesion may lead to systemic symptoms such as flushing and urticaria. Children with DCM present without distinct lesions, but rather a generalized thickening of skin (pachydermia) due to diffuse mast cell infiltration. DCM may be associated with bullae formation and more severe systemic symptoms including upper GI irritation and vascular collapse in the first few years of life. Maculopapular skin lesions of mastocytosis may be present in patients with adult-onset systemic disease. The apparent incidence of cutaneous lesions is $\geq 80\%$ in patients with ISM and $< 50\%$ in those with SM-AHNMD or ASM. In the upper gastrointestinal tract, gastritis and peptic ulcer are significant problems. In the lower intestinal tract, the occurrence of diarrhea and abdominal pain is attributed to increased motility due to mast cell mediators; this problem can be aggravated by malabsorption, which can also cause secondary nutritional insufficiency and osteomalacia. The periportal fibrosis associated with mast cell infiltration and a prominence of eosinophils may lead to portal hypertension and ascites. In some patients, anaphylaxis with rapid and life threatening vascular collapse can be induced by hymenoptera stings. These patients often have evidence of venom specific IgE. The neuropsychiatric disturbances are clinically most evident as impaired recent memory, decreased attention span, and "migraine-like" headaches. Patients may experience exacerbation of a specific clinical sign or symptom variably with alcohol ingestion, temperature changes, stress, use of mast cell-interactive opioids, or ingestion of NSAIDs.

DIAGNOSIS

Cutaneous mastocytosis is diagnosed by observing the characteristic lesions of MPCM or mastocytoma(s). A skin biopsy can be obtained to confirm these subvariants of CM, whereas patients with suspected DCM and bullous mastocytosis usually require a skin biopsy to confirm the diagnosis. Although the diagnosis of SM is generally suspected on the basis of the clinical history and physical findings, and can be supported by laboratory procedures, it can be established only by a tissue diagnosis. By convention, the diagnosis of SM depends heavily on bone marrow biopsy to meet the criteria of one major plus one minor or three minor findings (Table 347-3). The bone marrow provides the major criterion by revealing aggregates of mast cells, often in paratrabecular and perivascular locations with lymphocytes and eosinophils, as well as the minor criteria of abnormal mast cell morphology, aberrant mast cell membrane immunophenotype, or a codon 816 mutation in an extracutaneous tissue. A basal serum total tryptase level is a noninvasive approach to consider before bone marrow biopsy. The pro- β and α forms of tryptase are elevated in more than one-half of patients with SM and provide a minor criterion; the fully processed (“mature”) β form is increased in patients undergoing an anaphylactic reaction. A rare histopathologic subvariant called “well differentiated systemic mastocytosis” (WDSM) is characterized by clusters of mature appearing fully granulated and round mast cells, lack of aberrant CD25 and CD2 expression, and lack of D816V KIT mutation in most patients. These patients often have a history of childhood onset cutaneous disease and their mast cells may display aberrant CD30 expression and other markers of clonality such as atypical (non-D816V) KIT mutations. Additional studies directed by the presentation include a bone densitometry, bone scan, or skeletal survey; computed tomography scan, or endoscopy; and a neuropsychiatric evaluation. Osteoporosis is increased in mastocytosis and may lead to pathologic fractures.

TABLE 347-3

Diagnostic Criteria for Systemic Mastocytosis^a

Major: Multifocal dense infiltrates of mast cells (>15 mast cells per aggregate) in bone marrow or other extracutaneous tissues
Minor: Abnormal mast cell morphology (spindle shape, bi- or multi-lobed or eccentric nucleus, hypogranulated cytoplasm) Aberrant mast cell surface phenotype with expression of CD25 (IL-2 receptor alpha chain) and/or CD2 Detection of codon 816 mutation in peripheral blood cells, bone marrow cells, or an extracutaneous lesional tissue Total serum tryptase >20 ng/mL

^a
Diagnosis requires either the major criterion and one minor criterion or three minor criteria.

Some patients presenting with recurrent mast cell activation symptoms (particularly hypotensive syncopal anaphylactic episodes) have been found to have underlying mastocytosis. A subset of these patients may be found to have the D816V *KIT* mutation or aberrant mast cells displaying CD25, but lack other diagnostic criteria for SM. Such patients are termed to have “monoclonal mast cell activation syndrome.”

The differential diagnosis requires the exclusion of other flushing disorders. The 24-h urine assessment of 5-hydroxy-indoleacetic acid and metanephrines should exclude a carcinoid tumor or a pheochromocytoma, respectively. Some patients presenting with recurrent mast cell activation symptoms without an obvious increase in mast cell burden in skin or bone marrow have been shown to carry aberrant mast cells with clonality markers of D816V *KIT* mutation or surface CD25 expression. Most patients with recurrent IgE-induced or idiopathic anaphylaxis present with urticaria, angioedema, and/or wheezing, which are not manifestations of SM.

TREATMENT

TREATMENT

Mastocytosis

The management of SM uses a stepwise and symptom/sign-directed approach that includes an H₁ antihistamine for flushing and pruritus, an H₂ antihistamine or proton pump inhibitor for gastric acid hypersecretion, oral [cromolyn](#) sodium for diarrhea and abdominal pain, and occasionally aspirin (in those who are known to be tolerant of NSAIDs) for severe flushing with or without associated vascular collapse, despite use of H₁ and H₂ antihistamines, to block biosynthesis of PGD₂. Systemic glucocorticoids appear to alleviate the malabsorption. Mast cell cytoreductive therapy consisting of midostaurin, IFN-α or cladribine is generally reserved for advanced, nonindolent variants of SM. Midostaurin is a multi-kinase inhibitor with activity against D816V mutated and wild type KIT, and was recently approved by Food and Drug Administration for treatment of advanced systemic mastocytosis (SM-AHNMD, ASM, and MCL), and should be considered as a first line therapy of these disease variants. The efficacy of cytoreductive therapy in mastocytosis is variable, perhaps because of dosage limitations due to side effects. Imatinib is not effective in most cases as D816V *KIT* mutation provides resistance against it. Combination chemotherapy is appropriate for the frank leukemias. Stem cell transplantation has shown to be effective in a small subset of patients with advanced mastocytosis. A self-injectable [epinephrine](#) prescription is recommended for most patients due to increased incidence of anaphylaxis. Patients with a history of systemic hymenoptera venom reaction should be tested for venom specific IgE and placed on lifelong venom immunotherapy. Other investigational tyrosine kinase inhibitors with a capacity to inhibit D816V *KIT* mutation are currently in clinical trials.

FURTHER READING

Arock M et al: Current treatment options in patients with mastocytosis: Status in 2015 and future perspectives. Eur J Haematol 94:474, 2014.

Gotlib J et al: Efficacy and safety of midostaurin in advanced systemic mastocytosis. *N Engl J Med* 374:2530, 2016.

Hartmann K et al: Cutaneous manifestations in patients with mastocytosis: Consensus report of the European Competence Network on Mastocytosis; the American Academy of Allergy, Asthma & Immunology; and the European Academy of Allergology and Clinical Immunology. *J Allergy Clin Immunol* 137:35, 2016. [\[PubMed: 26476479\]](#)

Horny H-P et al: Mastocytosis (mast cell disease). In: *WHO Classification of Tumours. Pathology & Genetics. Tumours of Haematopoietic and Lymphoid Tissues*. SH Swerdlow, et al (eds). Lyon, France, IARC Press, 2008, pp 54–63.

Theoharides TC et al: Mast cells, mastocytosis, and related disorders. *N Engl J Med* 373:163, 2015. [\[PubMed: 26154789\]](#)

Ustun C et al: Consensus opinion on allogeneic hematopoietic cell transplantation in advanced systemic Mastocytosis. *Biol Blood Marrow Transplant* 22:1348, 2016. [\[PubMed: 27131865\]](#)

Valent P et al: Standards and standardization in mastocytosis: Consensus statements on diagnostics, treatment recommendations and response criteria. *Eur J Clin Invest* 37:435, 2007. [\[PubMed: 17537151\]](#)

Valent P et al: European Competence Network on Mastocytosis. Proposed diagnostic algorithm for patients with suspected mastocytosis: A proposal of the European Competence Network on Mastocytosis. *Allergy* 69:1267, 2014. [\[PubMed: 24836395\]](#)

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 353: Systemic Sclerosis (Scleroderma) and Related Disorders

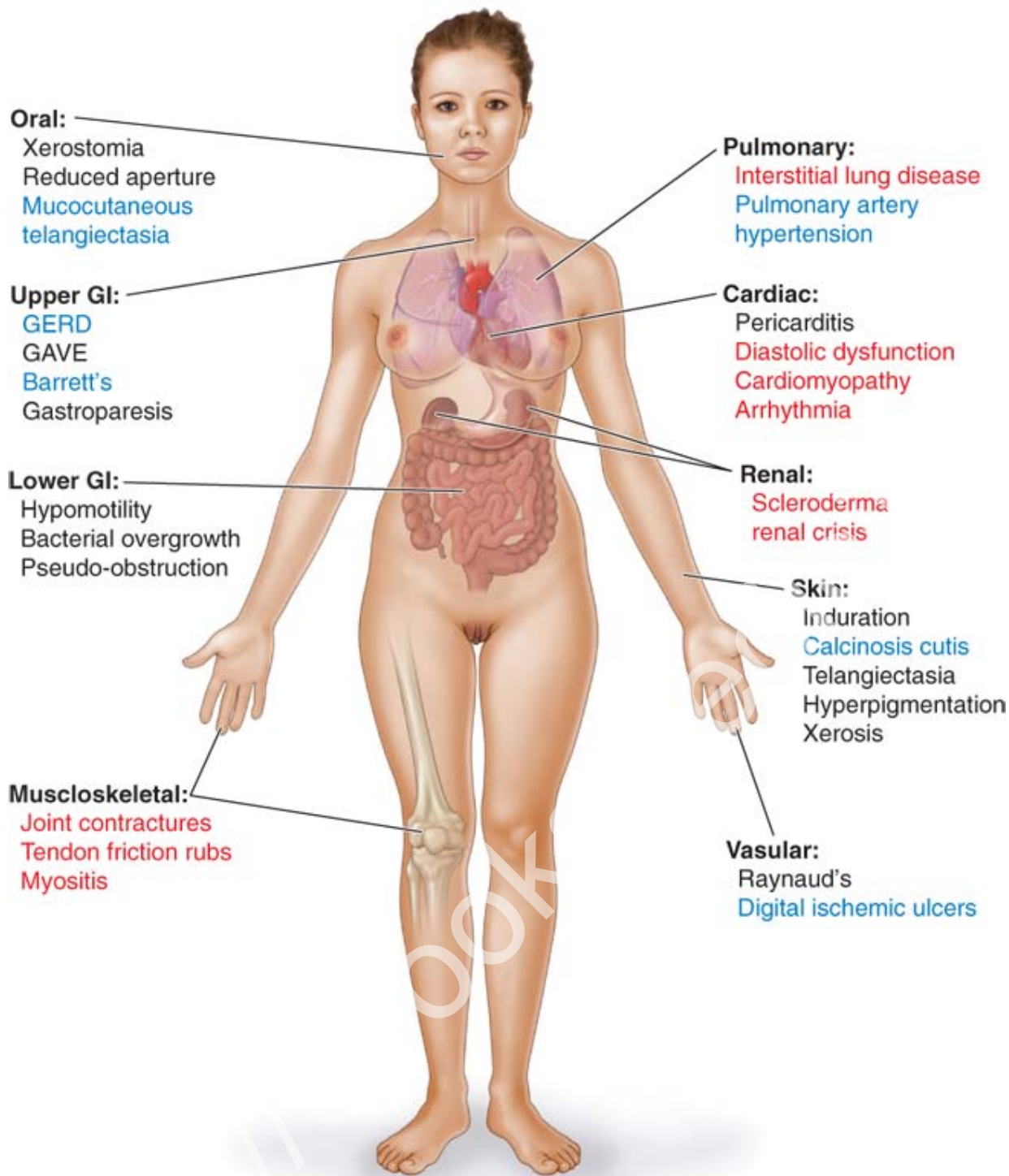
John Varga

DEFINITION AND CLASSIFICATION

Systemic sclerosis (SSc) is a complex and clinically heterogeneous orphan disease with protean clinical manifestations, a chronic and frequently progressive course, and significant disability, disfigurement and mortality. Virtually every organ can be affected ([Fig. 353-1](#)).

FIGURE 353-1

Multi-organ involvement in systemic sclerosis. Prominent complications more common in diffuse cutaneous SSc are shown in red; more common in limited cutaneous SSc in blue; and common in both forms of SSc shown in black.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

There is marked variability among SSc patients in patterns of skin involvement, organ complications, rates of disease progression, response to treatment, and survival. The early stages of SSc are associated with prominent inflammatory features; however, over time, structural alterations in multiple vascular beds and progressive visceral organ dysfunction due to fibrosis and atrophy come to dominate the clinical picture. Classification criteria for diagnosis of SSc are shown in [Table 353-1](#).

TABLE 353-1

Classification Criteria for Diagnosis of Systemic Sclerosis

ITEM	SUB-ITEM	WEIGHT/SCORE
Skin thickening (bilateral)—fingers extending proximal to MCP joints		9
Skin thickening of fingers only	Puffy fingers	2
	Sclerodactyly (skin thickened distal to MCP joints)	4
Fingertip lesions	Digital tip ulcer or pitting scar	2
		3
Mucocutaneous telangiectasia		2
Abnormal nails capillary pattern		2
Lung involvement	PAH	2
	Interstitial lung disease	2
Raynaud's phenomenon		3
SSc-specific autoantibodies	ACA	3
	Scl-70	
	RNA polymerase III	

Abbreviations: ACA, anterior cerebral artery; MCP, metacarpophalangeal joint; PAH, pulmonary arterial hypertension.

Although thick and indurated skin (*scleroderma*) is the distinguishing hallmark of SSc, skin changes also occur in localized forms of scleroderma, along with multiple metabolic, inherited and autoimmune disorders (**Table 353-2**). Patients with SSc can be broadly segregated into two major subsets defined by the pattern of skin involvement, clinical and laboratory features, and natural history (**Table 353-3**). Diffuse cutaneous SSc (dcSSc) is typically associated with extensive skin induration starting in the fingers (sclerodactyly) and ascending from distal to proximal limbs and the trunk. In these patients, interstitial lung disease (ILD) and acute renal involvement develop relatively early. In contrast, in patients with limited cutaneous SSc (lcSSc), Raynaud's phenomenon generally precedes other disease manifestations, sometimes by years. In these patients, skin involvement remains confined to the fingers, distal limbs, and face, while the trunk is spared. The constellation of calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia, was historically termed the *CREST syndrome*. In lcSSc, visceral organ involvement tends to show insidious progression, and digital ischemic ulcers, pulmonary arterial hypertension (PAH), hypothyroidism, and primary biliary cirrhosis may occur as late

complications. In some patients, Raynaud's phenomenon and characteristic clinical and laboratory features of SSc occur in the absence of detectable skin thickening. This syndrome has been termed *SSc sine scleroderma*.

TABLE 353-2

Conditions Associated with Skin Induration

Systemic sclerosis (SSc)

Limited cutaneous SSc

Diffuse cutaneous SSc

Localized scleroderma

Guttate (plaque) morphea, diffuse (pansclerotic) morphea, bullous morphea

Linear scleroderma, coup de sabre, hemifacial atrophy

Pansclerotic morphea

Overlap syndromes

Mixed connective tissue disease

SSc/polymyositis

Diabetic scleredema and scleredema of Buschke

Scleromyxedema (papular mucinosis)

Chronic graft-versus-host disease

Diffuse fasciitis with eosinophilia (Shulman's disease, eosinophilic fasciitis)

Stiff skin syndrome

Pachydermatoperiostosis (Primary hypertrophic osteoarthropathy)

Chemically induced and drug-associated scleroderma-like conditions

Vinyl chloride-induced disease

Eosinophilia-myalgia syndrome (associated with L-tryptophan contaminant exposure)

Nephrogenic systemic fibrosis (associated with gadolinium exposure)

Paraneoplastic syndrome

TABLE 353-3

Subsets of Systemic Sclerosis (SSc): Features of Limited Cutaneous SSc versus Diffuse Cutaneous Disease

CHARACTERISTIC FEATURE	LIMITED CUTANEOUS SSc	DIFFUSE CUTANEOUS SSc
Skin involvement	Indolent onset. Limited to fingers, distal to elbows, face; slow progression	Rapid onset. Diffuse: fingers, extremities, face, trunk; rapid progression
Raynaud's phenomenon	Antedates skin involvement, sometimes by years; may be associated with critical ischemia in the digits	Onset coincident with skin involvement; critical ischemia less common
Musculoskeletal	Mild arthralgia	Severe arthralgia, carpal tunnel syndrome, tendon friction rubs; small and large joint contractures
Interstitial lung disease	Slowly progressive, generally mild	Frequent, early onset and progression, can be severe
Pulmonary arterial hypertension	Frequent, late, may occur as an isolated complication	Often occurs in association with interstitial lung disease
Scleroderma renal crisis	Very rare	Occurs in 15%; onset may be fulminant; generally early (<4 years from disease onset)
Calcinosis cutis	Frequent, prominent	Less common, mild
Characteristic autoantibodies	Anti-centromere	Anti-topoisomerase I (Scl-70), anti-RNA polymerase III

INCIDENCE AND PREVALENCE

SSc is an acquired sporadic disease with a worldwide distribution and affecting all races. In the United States, the incidence is 9–46 cases per million per year. There are an estimated 100,000 U.S. cases, although this number may be significantly higher if patients who do not meet classification criteria are also included. There are large regional variations in incidence rates, potentially reflecting differences in case definition, environmental exposures or susceptibility genes in populations with different ancestries. Prevalence rates in England, Europe, and Japan appear to be lower than in North America and Australia. Age, sex, and ethnicity influence disease susceptibility, and blacks have higher age-specific incidence rates. In common with other connective tissue diseases, SSc shows a strong

female predominance (4.6:1), which is most pronounced in the childbearing years and declines after menopause. An additional risk factor is having an affected first-degree family member, which increases disease risk 13-fold. Although SSc can present at any age, the peak age of onset in women with both lcSSc and dcSSc is 65–74 years, although in blacks, disease onset occurs at an earlier age. Furthermore, blacks with SSc are more likely to have dcSSc, ILD, and a worse prognosis.

GENETIC CONTRIBUTION TO DISEASE PATHOGENESIS



SSc is a polygenic disease. In general, the genetic associations of SSc identified to date make only a small contribution to disease susceptibility. Disease concordance rates are low (4.7%) in monozygotic twins, although concordance for antinuclear antibody (ANA) positivity is significantly higher. On the other hand, evidence for genetic contribution to disease susceptibility is provided by the observation that 1.6% of SSc patients have a first-degree relative with SSc, a prevalence rate markedly increased compared to the general population. The risk of Raynaud's phenomenon, ILD, and other autoimmune diseases, including systemic lupus erythematosus (SLE) (**Chap. 349**), rheumatoid arthritis (**Chap. 351**), and autoimmune thyroiditis (**Chap. 375**), is also increased in first-degree relatives. Current approaches to uncover genetic factors in SSc include DNA sequencing and single nucleotide polymorphism (SNP) analysis of candidate genes, and SNP analysis of the entire genome in a hypothesis-free manner. Genome-wide association studies (GWASs) involve large multi-center and multi-national cohorts. A majority of the robustly validated susceptibility loci for SSc are genes involved in innate and adaptive immune responses, highlighting the importance of autoimmunity as the initial trigger for the disease. Genetic studies have shown associations with common (small effect size) variants related to B and T lymphocyte activation and signaling (*BANK1*, *BLK*, *CD247*, *STAT4*, *IL2RA*, *CCR6*, *IDO1*, *TNFSF4/OX40L*, *PTPN22*, and *TNIP1*). In addition, candidate gene studies and GWASs identified a strong association with human leukocyte antigen (HLA)-Class II haplotypes on chromosome 6, including *HLA-DRB1*11:04*, *DQA1*05:01*, and *DQB1*03:01*, and the non-HLA genes histocompatibility complex (MHC) genes *NOTCH4* and *PSORSC1*. Other genetic variants associated with SSc are involved in innate immunity and the interferon pathways (*IRF5*, *IRF7*, *STAT4*, *TNFAIP3/A20*, *GSDMA*, *PRDM1 (BLIMP1)*, *TNFAIP3*, and *TLR2*). Additional associations with *IL12RB2*, *IL-21*, the apoptosis-related genes *DNASE1L3* and *SOX5*, and the fibrosis-related genes *CSK*, *CAV1*, *PPARG*, and *GRB10* have been reported. In addition to disease susceptibility, some of these genetic loci are associated with particular disease manifestations or serologic subsets, including ILD (*CTGF*, *CD226*), PAH (*TNIP1*), and scleroderma renal crisis (*HLA-DRB1**). While the functional consequences of these gene variants and their potential roles in pathogenesis are currently not well understood, it seems likely that in combination they cause a state of altered immune regulation, leading to increased susceptibility to autoimmunity and persistent inflammation. Of note, many of the genetic variants associated with SSc are also implicated in other autoimmune disorders, including SLE, rheumatoid arthritis, and psoriasis, suggesting common pathogenic pathways shared among these phenotypically dissimilar conditions. The genetic associations identified to date only explain a fraction of the heritability of SSc, and GWASs, and whole exome sequencing to identify additional genetic susceptibility factors in SSc, particularly rare (and potentially causal) variants, are currently ongoing.

ENVIRONMENTAL AND OCCUPATIONAL RISK FACTORS

Given the relatively modest genetic contribution to disease susceptibility in SSc, environmental factors, such as infectious agents, intestinal microbiota, and occupational, dietary, lifestyle, and drug exposures, are likely to play a

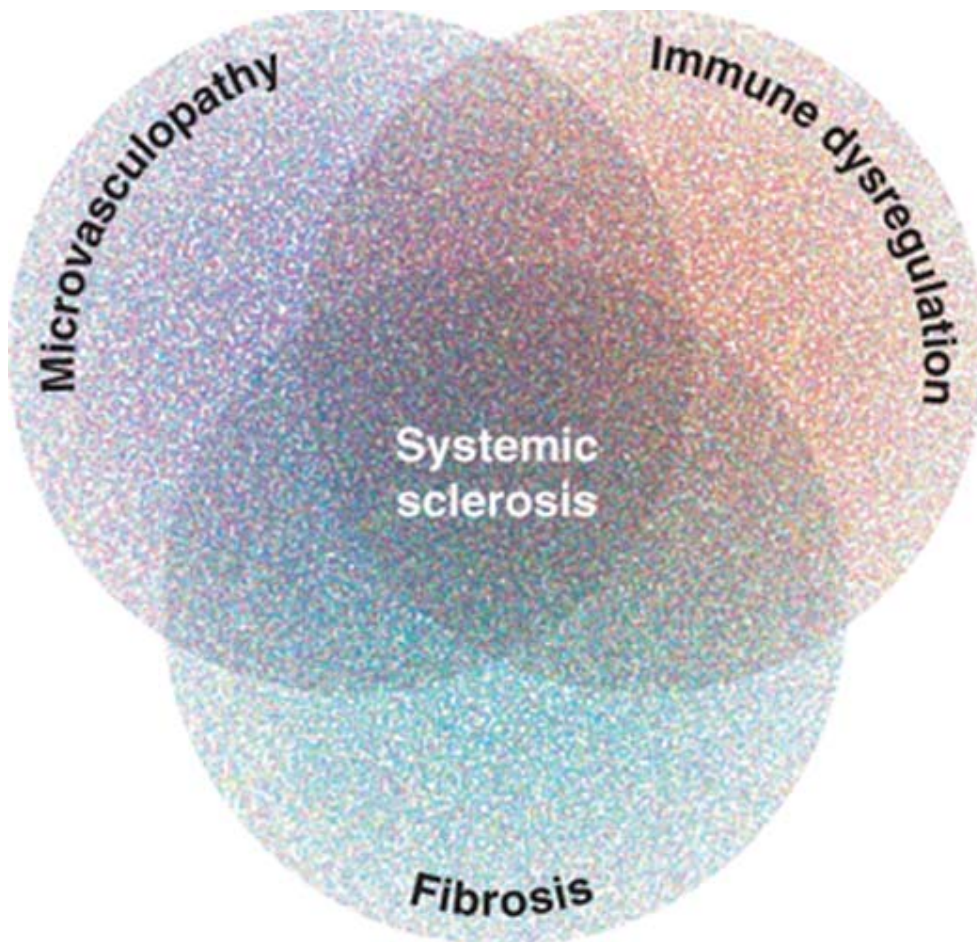
major role. Some evidence suggests potential roles for parvovirus B19, Epstein-Barr virus (EBV), cytomegalovirus (CMV), and *Rhodotorula glutinis* and other microorganisms. An epidemic of a novel syndrome with features suggestive of SSc occurred in Spain in the 1980s. The outbreak, termed *toxic oil syndrome*, was linked to use of contaminated rapeseed oil for cooking. Another epidemic outbreak, termed *eosinophilia-myalgia syndrome* (EMS), was linked to consumption of l-tryptophan-containing dietary supplements. Exposure to gadolinium contrast material in patients with renal compromise undergoing magnetic resonance scanning has been associated with nephrogenic systemic fibrosis. While each of these novel toxic-epidemic syndromes was characterized by chronic indurative skin changes and variable visceral organ involvement, the constellation of associated clinical, pathologic, and laboratory features distinguishes them from SSc. Occupational exposures tentatively linked with SSc include particulate silica (quartz), polyvinyl chloride, epoxy resins, welding fumes, and organic solvents and aromatic hydrocarbons including paint thinners, toluene, xylene, and trichloroethylene. These exposures might elicit stable and heritable epigenetic changes such as DNA methylation and histone modification underlying pathogenic alterations in gene expression. Drugs implicated in SSc-like illnesses include bleomycin, pentazocine, and cocaine, and appetite suppressants linked with PAH. Radiation therapy for cancer has been linked with *de novo* onset of SSc as well as with exacerbation of pre-existing SSc. In contrast to rheumatoid arthritis, cigarette smoking does not increase the risk of SSc. Although case reports and series of SSc in women with silicone breast implants had raised concern regarding a possible causal role of silicone in SSc, large-scale epidemiologic investigations found no evidence of increased prevalence of SSc.

PATHOGENESIS

Three cardinal pathomechanistic processes underlie the protean clinical manifestations of SSc: (1) diffuse microangiopathy, (2) inflammation and autoimmunity, and (3) visceral and vascular fibrosis in multiple organs (**Fig. 353-2**). While all three processes are concurrently operative in SSc patients, their activity, relative severity, and contribution to the overall clinical picture vary among individual patients and over time. In general, autoimmunity and altered vascular reactivity occur early, while fibrosis and atrophy occur later in the disease. Complex and dynamic interplay among these processes initiates and sustains the fibrotic process and tissue damage.

FIGURE 353-2

The characteristic constellation of vasculopathy, autoimmunity/inflammation and fibrosis underlies the protean clinical manifestations of systemic sclerosis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

ANIMAL MODELS OF DISEASE

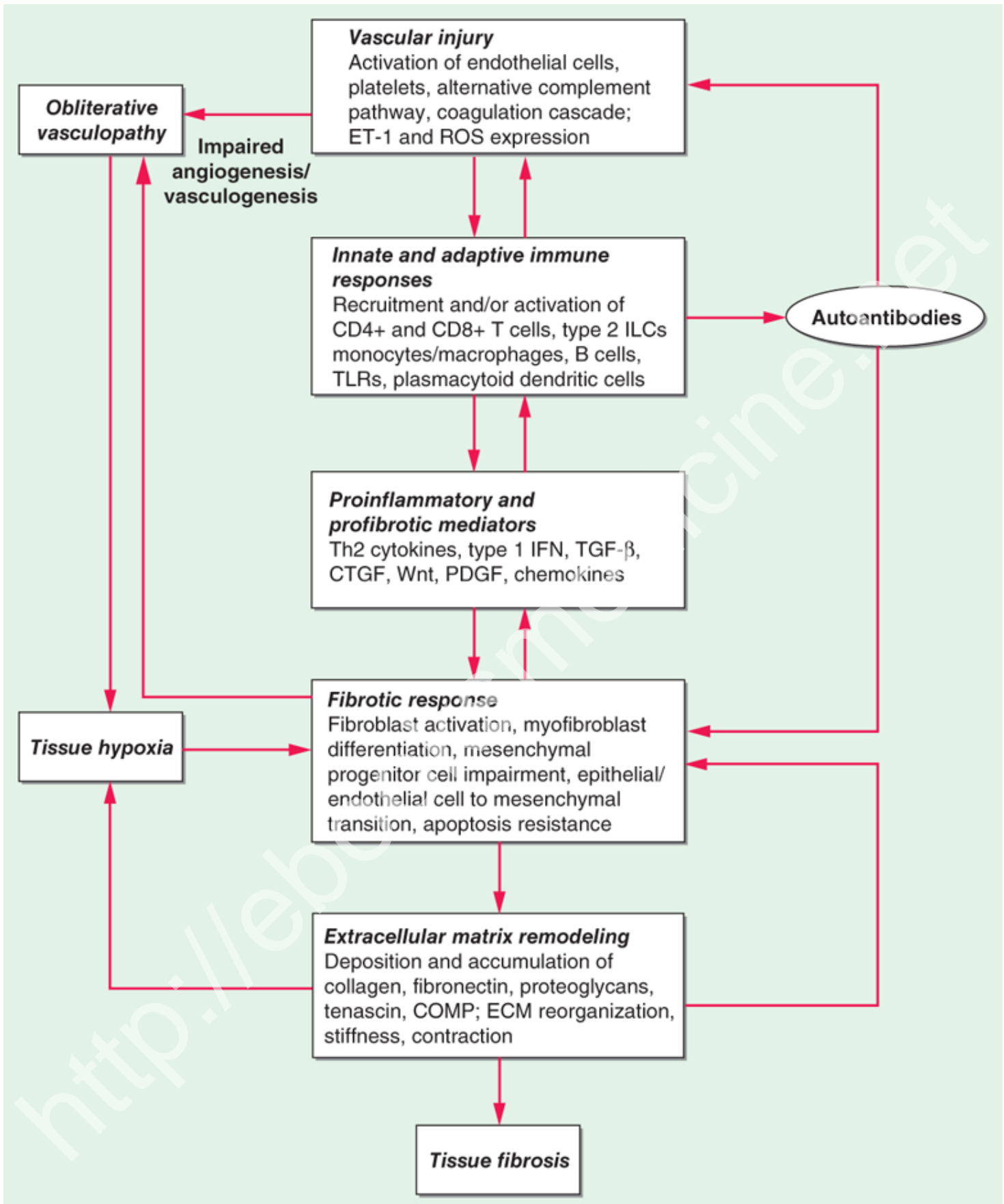
No single animal model of SSc fully reproduces the three cardinal processes that underlie pathogenesis, but some recapitulate selected aspects of the human disease. Tight-skin mice (Tsk1/+) spontaneously develop skin fibrosis due to a mutation in the fibrillin-1 gene. Mutant fibrillin-1 protein disrupts extracellular matrix assembly and causes aberrant activation of transforming growth factor β (TGF- β). Fibrillin-1 mutations in humans are associated with Marfan's disease and stiff skin syndrome, but have not been reported in SSc. Skin and lung fibrosis accompanied by variable vasculopathy and autoimmunity can be elicited in mice by injection of bleomycin or Angiotensin II, or by transplantation of HLA-mismatched bone marrow or spleen cells. Targeted genetic modifications in mice give rise to new disease models for investigating the pathogenetic roles of individual molecules, pathways, and cell types. For example, mice lacking IRF5, Smad3, uPAR, or peroxisome proliferator-activated receptor (PPAR)- γ , or constitutively overexpressing β -catenin, Wnt10b, sirtuin 3, Fra-2, TGFB1, PDGFR α , or adiponectin are either resistant or hypersensitive to experimental scleroderma, or spontaneously develop fibrosis. These disease models can contribute to understanding specific aspects of SSc pathogenesis, and to discovery and validation of novel targets for therapy.

MICROANGIOPATHY

Vascular injury is an early and possibly primary pathogenic event in SSc that leads to protean clinical manifestations of small vessel vasculopathy ([Fig. 353-3](#)).

FIGURE 353-3

Initial vascular injury in a genetically susceptible individual triggers functional and structural vascular alterations, inflammation and autoimmunity, culminating in fibrosis. Inflammatory and immune responses initiate and sustain fibroblast activation and differentiation, resulting in pathologic fibrogenesis and irreversible tissue damage. Vascular damage results in tissue ischemia that further contributes to progressive fibrosis and atrophy. COMP, cartilage oligomeric matrix protein; CTGF, connective tissue growth factor; PDGF, platelet-derived growth factor; ROS, reactive oxygen species; TGF- β , transforming growth factor β ; TLR, toll-like receptor.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Prominent microangiopathy in multiple vascular beds has important clinical sequelae including mucocutaneous telangiectasiae, Raynaud's phenomenon, ischemic digital ulcers, scleroderma renal crisis, myocardial involvement, and PAH. Raynaud's phenomenon is characterized by altered blood-flow response to cold challenge in small digital arteries. This initially reversible functional abnormality is associated with autonomic and peripheral nervous system alterations, including impaired production of the neuropeptide calcitonin gene-related peptide from sensory afferent nerves and heightened sensitivity of α_2 -adrenergic receptors on vascular smooth-muscle cells. Isolated (primary) Raynaud's disease is common, generally benign and non-progressive. In contrast, secondary Raynaud's phenomenon in SSc is often progressive and complicated by irreversible structural changes, culminating in ischemic digital ulcers, necrosis, and amputation.

Viruses, cytotoxic factors, and chemokines thrombogenic microparticles, alternate complement pathway activation and autoantibodies targeting endothelial cells, phospholipids, and β_2 glycoprotein I (β_2 GPI) are implicated as potential triggers of endothelial cell injury. Endothelial damage results in dysregulated production of vasodilatory (nitric oxide and prostacyclin) and vasoconstricting (endothelin-1) substances, as well as upregulation of intercellular adhesion molecule 1 (ICAM-1) and other surface adhesion molecules. Microvessels show enhanced permeability and transendothelial leukocyte diapedesis, abnormal activation of coagulation cascades, elevated **thrombin** production, and impaired fibrinolysis. Spontaneous platelet aggregation causes release of serotonin, platelet-derived growth factor (PDGF), and platelet alpha granules including thromboxane, a potent vasoconstrictor. Smooth-muscle cell-like myointimal cells in the media proliferate, the basement membrane is thickened and reduplicated, and perivascular adventitial fibrosis develops. The vasculopathic process affects capillaries, as well as arterioles, and less commonly even large vessels in many organs, resulting in reduced blood flow and tissue ischemia. Progressive luminal occlusion due to intimal and medial hypertrophy, combined with persistent endothelial cell damage and adventitial fibrosis, establish a vicious cycle that culminates in the striking absence of small blood vessels (rarefaction) in late-stage disease. Recurrent ischemia-reperfusion generates reactive oxygen species (ROS) that further damages the endothelium through peroxidation of membrane lipids. Paradoxically, the process of revascularization that normally reestablishes blood flow to ischemic tissue is defective in SSc despite elevated levels of other angiogenic factors. Moreover, bone marrow-derived circulating endothelial progenitor cells are reduced in number and impaired in function. Widespread capillary loss, obliterative vasculopathy of small and medium-sized arteries, and impaired ability to repair and replace damaged vessels are hallmarks of SSc.

INFLAMMATION AND AUTOIMMUNITY

Cellular Immunity

The following observations provide support for the inflammatory/autoimmune nature of SSc: near-universal presence of circulating autoantibodies with defined specificities; familial clustering of SSc with other autoimmune diseases; detection of activated immune cells, including T cells with oligoclonal antigen receptors, in target organs; prominent type I interferon (IFN) signatures, characterized by elevated expression of IFN-regulated genes, in a variety of cell types; elevated circulating levels and spontaneous secretion from mononuclear cells of cytokines and chemokines such as interleukin-6 (IL-6); tumor necrosis factor, IL-4, IL-10, IL-17, IL-33, CCL2, and CXCL4; genetic association of SSc with variants of MHC and other genes functionally implicated in the immune response; and the rapid clinical response, fibrosis resolution, and vascular regeneration observed in some SSc patients treated with immunomodulatory or immunoablative therapies. Genetic studies reveal strong associations with MHC locus

alleles, as well as non-HLA-linked genes encoding mediators of both adaptive and innate immune responses (*CD247*, *STAT4*, *IRF5*, *CD226*, *TNFAIP3/A20*, and *TNFSF4*).

Circulating monocytes from SSc patients overexpress IFN-regulated genes such as Siglec-1, have reduced levels of caveolin-1, and exhibit an inherently profibrotic phenotype. In early (edematous) stage SSc, mononuclear cell infiltrates comprised of activated T cells, monocytes/macrophages, and dendritic cells can be seen in skin, lungs, and other affected organs prior to appearance of fibrosis or vascular damage. Dendritic cells can be found in close proximity to activated fibroblasts and myofibroblasts and express toll-like receptors (TLR) and secrete IFN, IL-10, thymic stromal lymphopoietin (TSLP), and CXCL4, shaping the adaptive immune response and contributing to loss of immune tolerance. Tissue-infiltrating T cells express CD45 and HLA-DR activation markers and display restricted T cell receptor signatures indicative of oligoclonal expansion in response to recognition of as-yet unknown antigen. Of note, in patients diagnosed with SSc in close temporal association with cancer who are RNA polymerase III antibody-positive, the tumor may show mutations in RNAPol3 autoantigen, which results in the generation of mutant-specific T cell immunity and cross-reactive antibodies. These findings support the premise that an abnormal antigen might act as initial trigger for the autoimmune response in SSc.

Circulating T cells in SSc express chemokine receptors and α_1 integrin, accounting for their enhanced binding to endothelium and to fibroblasts, while endothelial cells express ICAM-1 and other adhesion molecules that facilitate leukocyte diapedesis. Activated T cells show a T_H2 -polarized immune response driven by dendritic cells. The T_H2 cytokines IL-4, IL-13, IL-33, and TSLP induce fibroblast activation, whereas the T_H1 cytokine interferon γ (IFN- γ) blocks cytokine-mediated fibroblast activation and exhibits anti-fibrotic properties. Evidence for altered Th17 and regulatory T cell (Treg) numbers and function in SSc has been reported. Type 2 innate lymphoid cells (iLCs), a recently discovered lymphoid cell population implicated in type 2 immunity and tissue remodeling, are also elevated in SSc skin biopsies. Alternately activated M2 macrophages, which produce TGF- β and promote angiogenesis and tissue remodeling, are increased in the skin in SSc. Although the frequency of regulatory T cells that enforce immune tolerance is elevated in the circulation and tissues, their immunosuppressive function appears to be defective. Some evidence implicates altered B cell homeostasis and function in SSc. Circulating B cells show elevated CD19 and co-stimulatory molecules CD80 and CD86, suggesting B cell chronic activation. Serum levels of a proliferation-inducing ligand (APRIL) and B cell activating factor (BAFF), members of the TNF superfamily with potent effects on B cell activation, are elevated in SSc, and associate with extent of skin and lung involvement. B cells secrete IL-6, TGF- β , and other profibrotic cytokines implicated in pathogenesis. Thus, B cell hyperactivity in SSc might directly contribute to the inflammatory and fibrotic processes, as well as generation of autoantibodies. Microarray analysis identifies a distinct subset of SSc skin biopsies with elevated expression of inflammation-related genes. Evidence of innate immune and TLR signaling, reflecting activation by type 1 IFN from plasmacytoid dendritic cells, is prominent in peripheral blood cells and target organs.

Humoral Autoimmunity

Circulating ANAs can be detected by indirect immunofluorescence in virtually all patients with SSc, even in early stages of disease. In addition, several SSc-specific autoantibodies with distinct patterns of immunofluorescence show strong associations with unique disease endophenotypes (**Table 353-4**). These antibodies are directed mostly against intracellular proteins associated with transcription, DNA repair, and RNA processing. Owing to their high specificity, mutual exclusivity and association with unique disease manifestations, SSc-associated autoantibodies

have substantial utility in clinical practice as diagnostic and prognostic markers, while their role in monitoring disease activity remains uncertain. Moreover, antibodies directed against fibrillin-1, matrix metalloproteinases, cell surface markers Angiotensin II receptor, endothelin-1 receptor, muscarinic 3 receptor, or the PDGF receptor, have been described in patients with SSc, although their clinical relevance is not yet established. These antibodies manifest functional receptor agonist activity and might have direct pathogenic roles.

TABLE 353-4

Major Systemic Sclerosis-Specific Autoantibodies and Principal Associated Features

TARGET ANTIGEN	SSc SUBSET	PROMINENT CHARACTERISTIC CLINICAL ASSOCIATION
DNA Topoisomerase I (Scl-70) Speckled pattern	dcSSc	Tendon friction rubs, digital ischemic ulcers, scleroderma, extensive skin involvement, early ILD, cardiac involvement, scleroderma renal crisis
Centromere proteins Discrete speckled (centromere) pattern	lcSSc	Digital ischemic ulcers, calcinosis cutis, isolated PAH; renal crisis rare
RNA polymerase III Speckled pattern	dcSSc	Rapidly progressive skin, tendon friction rubs, joint contractures, GAVE, renal crisis, contemporaneous cancers; digital ulcers rare
U3-RNP (fibrillarin) Nucleolar pattern	dc/lcSSc	PAH, ILD, scleroderma renal crisis, GI tract involvement, myositis
Th/T ₀ Nucleolar pattern	lcSSc	ILD, PAH
PM/Scl Nucleolar pattern	lcSSc	Calcinosis cutis, ILD, myositis overlap
Ku Speckled pattern	Overlap	SLE, myositis overlap
U1-RNP Speckled pattern	MCTD	PAH, inflammatory arthritis, myositis overlap
U11/U12 RNP Speckled pattern	dc/lcSSc	ILD

Abbreviations: dcSSc, diffuse cutaneous SSc; GAVE, gastric antral vascular ectasia; ILD, interstitial lung disease; lcSSc, limited cutaneous SSc; MCTD, mixed connective tissue disease; PAH, pulmonary arterial hypertension; SLE, systemic lupus erythematosus.

A variety of mechanisms have been proposed to account for the generation of SSc-associated autoantibodies. Proteolytic cleavage, increased expression or altered subcellular localization of normal proteins, or their alterations due to mutation in the case of certain tumors, could lead to immune recognition as neoepitopes, resulting in the breaking of immune tolerance.

FIBROSIS

Fibrosis affecting multiple organs is a distinguishing feature of SSc. The process is characterized by replacement of normal tissue architecture with dense, rigid, avascular, and relatively acellular connective tissue. Fibrosis in SSc follows, and is a consequence of, inflammation, autoimmunity, and microvascular damage (Fig. 353-3). Fibroblasts are mesenchymal cells primarily responsible for the functional and structural integrity of connective tissue. Upon their activation by extracellular cues, fibroblasts proliferate, migrate, secrete collagens and other matrix molecules, growth factors, chemokines, and cytokines, and transdifferentiate into contractile myofibroblasts. Under normal conditions, these self-limited responses accomplish physiologic repair and regeneration of tissue. In contrast, when these responses become sustained and amplified, pathologic fibrosis results. Stimulatory signaling by endogenous TGF- β and paracrine fibrotic mediators including IL-6, IL-13, Wnt ligands, connective tissue growth factor (CTGF), PDGF, lysophosphatidic acid, endothelin-1, hypoxia, ROS, **thrombin**, and mechanical forces are responsible for sustained fibroblast activation underlying non-resolving fibrosis in SSc. Buildup of damage-associated endogenous ligands for TLR4 (EDA-fibronectin, high mobility group B1 (HMGB1) and Tenascin-C) and for TLR9 (mitochondrial DNA) within the fibrotic microenvironment further contributes to non-resolving fibrosis.

In addition to tissue-resident fibroblasts and transformed myofibroblasts, bone marrow–derived circulating mesenchymal progenitor cells also contribute to fibrosis. The factors that regulate the differentiation of mesenchymal progenitor cells and their trafficking from the circulation into lesional tissue are unknown. Epithelial and endothelial cells, mesenchymal progenitor cells, preadipocytes and tissue fibroblasts have all been proposed as sources of myofibroblasts in fibrosis. Although myofibroblasts are transiently found in normal wound healing, their persistence in fibrotic tissue, possibly due to resistance to apoptosis, contributes to scar formation.

Explanted SSc fibroblasts display an abnormally activated phenotype *ex vivo*, with variably increased rates of collagen production, spontaneous ROS generation, prominent stress fibers, and constitutive expression of alpha smooth-muscle actin. Persistence of the “scleroderma phenotype” during serial *ex vivo* passage of SSc fibroblasts may reflect autocrine TGF- β stimulatory loops, deregulated microRNA expressions, or stable acquired epigenetic modifications in these cells.

PATHOLOGY

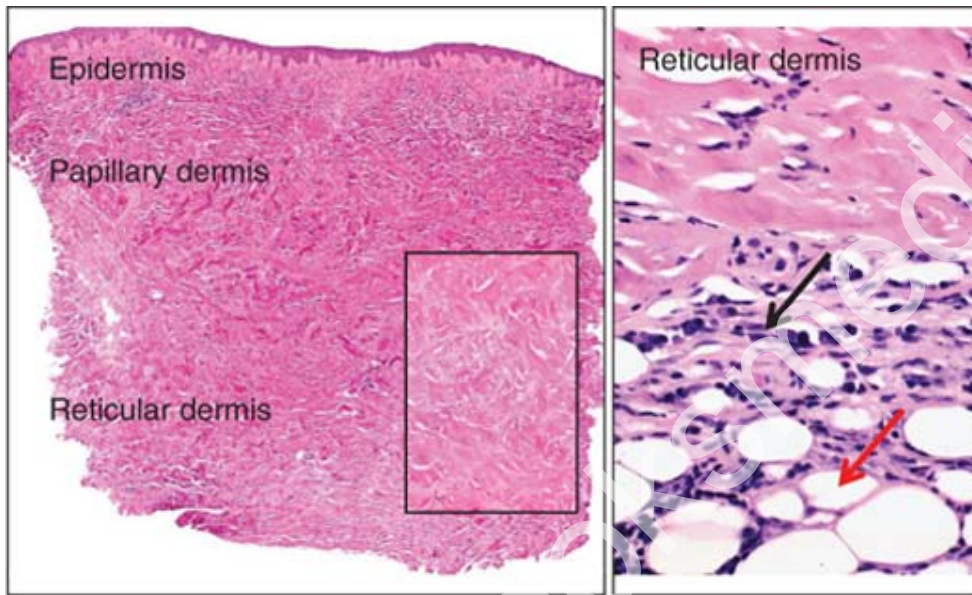
While pathological findings in SSc vary across anatomic sites, the distinguishing hallmark of SSc irrespective of the organ system is the triad of widespread capillary loss and obliterative microangiopathy, combined with fibrosis in the skin and internal organs. In early-stage disease, perivascular inflammatory cell infiltrates composed of T and B lymphocytes, activated monocytes and macrophages and mast cells may be detected in multiple organs. A non-inflammatory obliterative microangiopathy is a prominent late finding in the heart, lungs, kidneys, and gastrointestinal tract. Fibrosis is found in the skin, lungs, cardiovascular and gastrointestinal systems, tendon sheaths, perifascicular tissue surrounding skeletal muscle, and some endocrine organs. Excessive accumulation of collagens, proteoglycans, COMP and other structural matrix macromolecules progressively disrupts normal architecture, resulting in impaired function and failure of affected organs.

SKIN

The dermis is thickened, and accumulation of broad bundles of homogenized collagen oriented parallel to the epithelium is seen (**Fig. 353-4A**). Adnexal glands are atrophic, and loss of periadnexal and intradermal white adipose tissue and its replacement with collagen can be striking. While perivascular mononuclear cell infiltrates may be seen early, established skin fibrosis generally shows absence of inflammation. These findings are histologically indistinguishable from those in localized scleroderma.

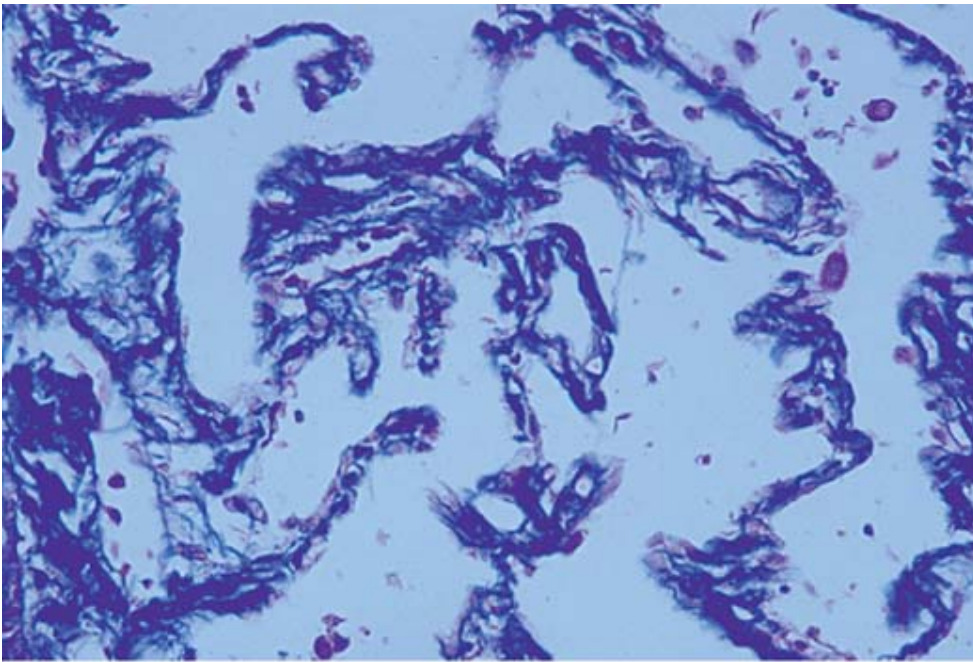
FIGURE 353-4

Pathologic findings in systemic sclerosis (SSc). **A. Left panel:** The skin is thickened due to expansion of the dermis. Inset, higher magnification showing thick hyalinized collagen bundles replacing skin appendages. **Right panel:** Mononuclear inflammatory cells within the intradermal adipose tissue. Black arrow, collagen; red arrow, dermal adipocytes. **B.** Early SSc-ILD. Diffuse fibrosis of the alveolar septae and a chronic inflammatory cell infiltrate. Trichrome stain. **C.** Pulmonary arterial obliterative vasculopathy. Striking intimal hyperplasia and luminal narrowing of small pulmonary artery, with little inflammation and minimal interstitial fibrosis, in a patient with SSc-PAH.

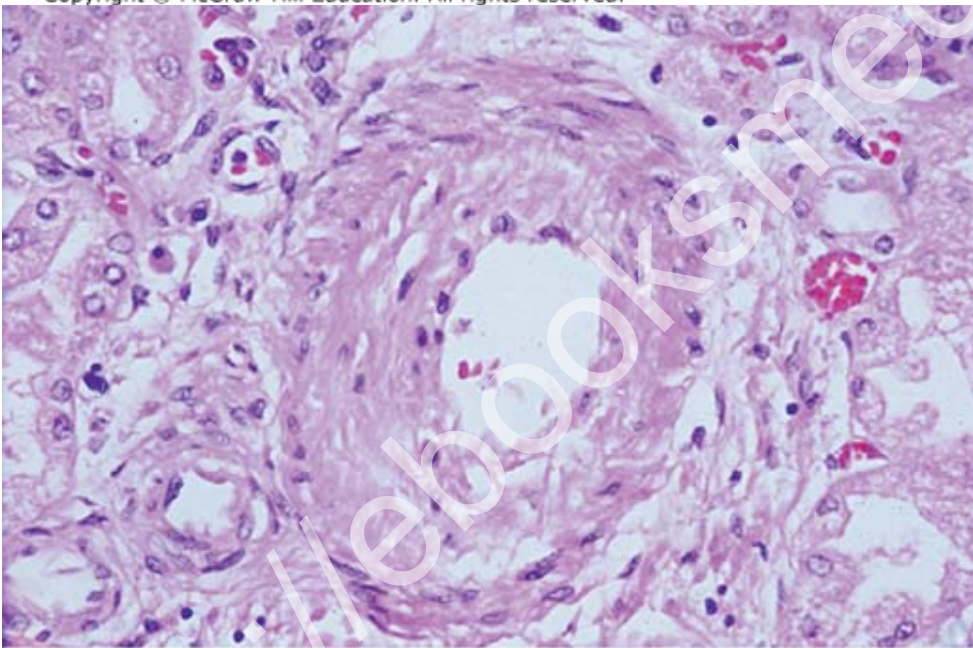


A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

**B**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

**C**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

LUNGS

Autopsy studies in SSc universally show evidence of lung involvement. Most common is a nonspecific interstitial pneumonia (NSIP) pattern characterized by variable interstitial fibrosis and mild chronic inflammation. Patchy infiltration of the alveolar walls with T lymphocytes, macrophages, and eosinophils may occur in early disease. With progression, interstitial fibrosis and vascular damage dominate, often coexisting within the same biopsy. The usual interstitial pneumonia (UIP) pattern of spatial/temporal heterogeneity of inflammation, fibrosis and fibrotic foci

seen in idiopathic pulmonary fibrosis is less common in SSc (**Fig. 353-4B**). Fibrosis of the alveolar septae results in obliteration of the airspaces and loss of pulmonary blood vessels. This process impairs gas exchange and contributes to pulmonary hypertension. Intimal thickening of the pulmonary arteries, best seen with elastin stain, underlies SSc-associated PAH (**Fig. 353-4C**) and, at autopsy, is often associated with multiple pulmonary emboli and myocardial fibrosis. Patients may also show fibrosis and intimal proliferation in preseptal venules and veins in the lung, accounting for veno-occlusive disease. Lymphocytic bronchiolitis involving the submucosa of the terminal bronchioles may also be seen.

GASTROINTESTINAL TRACT

Pathologic changes can be found at any level from the mouth to the rectum. Atrophy and fibrosis of the muscularis propria and characteristic vascular lesions are prominent in the lower esophagus, while striated muscle in the upper third of the esophagus is generally spared. Collagenous replacement of the normal intestinal tract architecture results in impaired smooth muscle contractility and diminished peristaltic activity, with dysmotility, bacterial overgrowth, small-bowel obstruction, and perforation. Chronic gastroesophageal reflux is associated with esophageal inflammation, mucosal ulceration, and stricture formation and may lead to Barrett's metaplasia with attendant risk of adenocarcinoma. Esophageal dilation and reflux are associated with ILD due to chronic microaspiration.

KIDNEYS

In the kidneys, vascular lesions affecting the interlobular and arcuate arteries predominate. Chronic renal ischemia is associated with shrunken glomeruli. Patients with scleroderma renal crisis show acute fibrinoid necrosis of afferent arterioles, followed by intimal proliferation (onion-skin pattern), and ischemic collapse of glomeruli. These changes are reminiscent of thrombotic microangiopathies such as atypical hemolytic-uremic syndrome (**see Chap. 304**), and are accompanied by complement deposition, thrombosis, thrombocytopenia due to platelet consumption, and intravascular hemolysis. Extensive vascular thrombosis, glomerular collapse and sclerosis, and peritubular capillary deposits in renal biopsy are associated with irreversible renal failure.

HEART

Subclinical cardiac pathology is common, with prominent involvement of the myocardium and pericardium. The characteristic arteriolar lesions are concentric intimal hypertrophy and luminal narrowing, accompanied by patchy contraction band necrosis, loss of cardiac myocytes, and myocardial fibrosis due to microvascular involvement and ischemia-reperfusion injury. Fibrosis of the conduction system is common, especially at the sinoatrial node. The frequency of epicardial atherosclerotic coronary artery disease may be increased compared to the general population, similar to other systemic inflammatory diseases. Pericardial involvement with chronic inflammatory infiltrates and fibrinous exudates is common and may be associated with pericardial effusions.

PATHOLOGY IN OTHER ORGANS

Synovitis may be found in early SSc; with disease progression, the synovium becomes fibrotic, and in contrast to rheumatoid disease, pannus formation or bone resorption are uncommon. Fibrosis of tendon sheaths and fascia, sometimes accompanied by calcifications, produces palpable and sometimes audible tendon friction rubs.

Inflammation and, in later stages, atrophy and fibrosis of skeletal muscles are common findings, and are similar to those in polymyositis. Fibrosis of the thyroid gland and of the minor salivary glands may be seen. Placentas from SSc pregnancies show decidual vasculopathy, which is associated with poor perinatal outcomes and fetal death.

CLINICAL FEATURES

OVERVIEW

SSc can affect virtually any organ ([Fig. 353-1](#) and [Table 353-5](#)). Although a dichotomous approach stratifying SSc into diffuse and limited cutaneous subsets ([Table 353-2](#)) is useful, disease expression is far more complex, and multiple distinct endophenotypes with unique patterns of manifestations can be recognized within each subset. Unique endophenotypes associate with autoantibodies with distinct and mutually exclusive specificities ([Table 353-4](#)). Patients with SSc “overlap” have typical features coexisting with clinical and laboratory evidence of another autoimmune disease, most commonly polymyositis, Sjögren’s syndrome, polyarthritis, autoimmune liver disease, or SLE.

TABLE 353-5

Frequency of Clinical Organ Involvement in Limited Cutaneous and Diffuse Cutaneous Systemic Sclerosis (SSc)

FEATURES	LIMITED CUTANEOUS SSc (%)	DIFFUSE CUTANEOUS SSc (%)
Skin involvement	90 ^a	100
Raynaud’s phenomenon	99	98
Ischemic digital ulcers	50	25
Esophageal involvement	90	80
Interstitial lung disease	35	65
Pulmonary arterial hypertension	15	15
Myopathy	11	23
Clinical cardiac involvement	9	12
Scleroderma renal crisis	2	15
Calcinosis cutis	—	—

^aApproximately 10% of patients have SSc *sine* scleroderma.

INITIAL CLINICAL PRESENTATION

Characteristic initial presentation is quite different in patients with the diffuse (dcSSc) versus limited (lcSSc) cutaneous forms of the disease. In dcSSc, the interval between Raynaud's phenomenon and onset of other disease manifestations is brief (weeks to months). Soft tissue swelling, puffy fingers, and intense pruritus are signs of the early inflammatory "edematous" phase. The fingers, distal limbs, and face are usually affected first. Diffuse hyperpigmentation of the skin, carpal tunnel syndrome arthralgias, muscle weakness, fatigue, and decreased joint mobility are common. During the ensuing weeks to months, the inflammatory edematous phase evolves into the "fibrotic" phase, with skin induration associated with hair loss, reduced production of skin oils, and decline in sweating capacity. Progressive flexion contractures of the fingers ensue. The wrists, elbows, shoulders, hip girdles, knees, and ankles become stiff due to fibrosis of the supporting joint structures. While advancing skin involvement is the most visible manifestation of early dcSSc, important and clinically silent internal organ involvement commonly occurs during this stage. The initial 4 years from disease onset is the period of most rapidly evolving pulmonary and renal damage. If organ failure does not occur during this phase of dcSSc, the systemic process may stabilize.

Compared to dcSSc, the course of lcSSc tends to be more indolent. The interval between onset of Raynaud's phenomenon and disease manifestations such as GERD, cutaneous telangiectasia, or soft tissue calcifications can be as long as years. Scleroderma renal crisis, significant ILD, and tendon friction rubs occur rarely in lcSSc, while PAH, and overlap with keratoconjunctivitis sicca, polyarthrititis, cutaneous vasculitis, and biliary cirrhosis can develop many years after disease onset.

ORGAN INVOLVEMENT

RAYNAUD'S PHENOMENON

Raynaud's phenomenon, the most frequent extracutaneous complication of SSc, is characterized by episodes of reversible vasoconstriction in the fingers and toes, sometimes also affecting the tip of the nose and earlobes. Attacks, triggered by a decrease in temperature, as well as emotional stress and vibration, typically start with pallor, followed by cyanosis of variable duration. Hyperemia ensues spontaneously or with rewarming of the digit. The progression of the three color phases reflects the underlying vasoconstriction, ischemia, and reperfusion. Up to 5% of the general population has Raynaud's phenomenon. In the absence of signs or symptoms of an underlying condition, Raynaud's phenomenon is classified as primary (Raynaud's disease), which represents an exaggerated physiologic response to cold. Secondary Raynaud's phenomenon occurs in SSc and other connective tissue diseases, hematologic and endocrine conditions, and occupational disorders, and can complicate treatment with beta blockers and anticancer drugs such as cisplatin and bleomycin. Distinguishing primary Raynaud's disease from secondary Raynaud's phenomenon can present a diagnostic challenge. Raynaud's disease is supported by the following: absence of an underlying cause, a family history of Raynaud's phenomenon, absence of digital tissue necrosis or ulceration, and a negative ANA test. Secondary Raynaud's phenomenon tends to occur at an older age (>30 years), is more severe (episodes more frequent, prolonged, and painful), and is associated with ischemic digital ulcers and loss of digits ([Fig. 353-5](#)).

FIGURE 353-5

Digital necrosis. Sharply demarcated necrosis of the fingertip secondary to ischemia in a patient with limited cutaneous systemic sclerosis (SSc) associated with severe Raynaud's phenomenon.

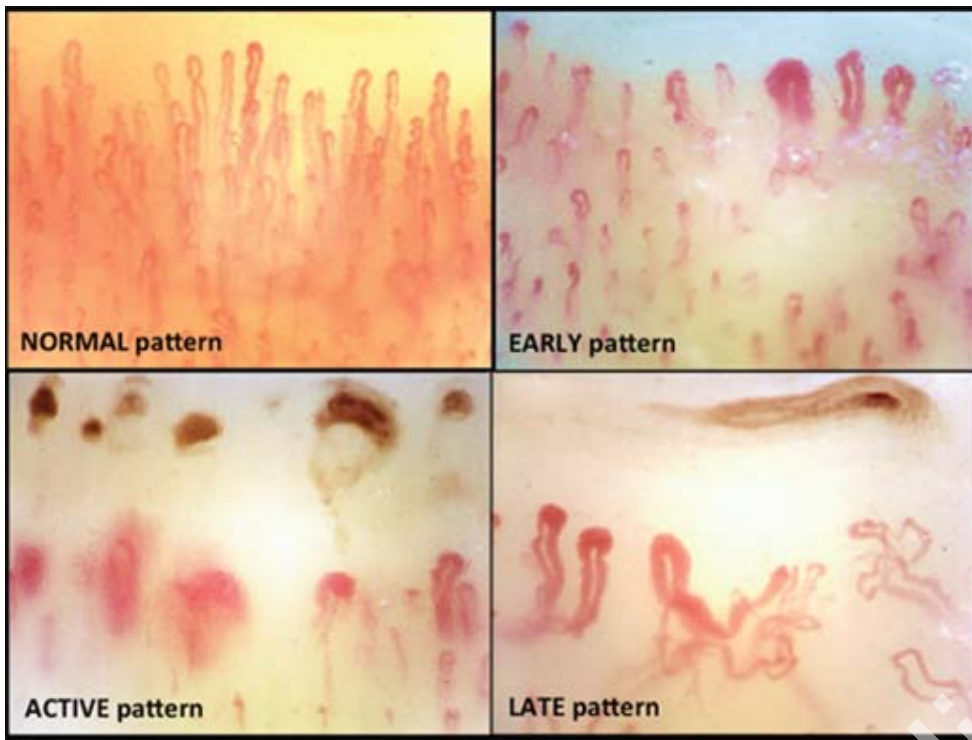


Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Nailfold capillaroscopy using a low-power stereoscopic microscope or ophthalmoscope permits visualization of nailbed cutaneous capillaries under immersion oil (**Fig. 353-6**). Raynaud's disease is associated with evenly spaced parallel vascular loops, whereas in secondary Raynaud's phenomenon, nailfold capillaries are distorted with widened and irregular loops, dilated lumen, microhemorrhages, and areas of vascular "dropout." Thus, nailfold capillaroscopy can be helpful in differentiating primary from secondary Raynaud's phenomenon and in establishing the early diagnosis of SSc.

FIGURE 353-6

SSc-associated nailfold capillary alterations. Normal nailfold pattern in healthy subjects. Note regularly-arrayed and uniform-size "hairpin" microvessels; "early pattern" showing dilations of microvessels and symmetrically increased microvessels (giant capillaries) representing the first morphological sign of systemic sclerosis; "active pattern" with giant capillaries, collapse with microhemorrhages and loss of capillaries; "late pattern" showing massive loss of capillaries, fibrosis (white/yellow background) and neoangiogenesis with secondary dilations (nailfold videocapillaroscope VIDEOCAP; magnification 220×). (Courtesy of Professor Maurizio Cutolo, University of Genoa.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

SKIN FEATURES

Bilateral symmetrical skin thickening is the hallmark of SSc that distinguishes it from other connective tissue diseases. Skin involvement starts in the fingers and characteristically advances from distal to proximal extremities in an ascending fashion. Some patients note diffuse tanning in the absence of sun exposure as a very early manifestation. In dark-skinned individuals, vitiligo-like hypopigmentation may occur. Because pigment loss spares the perifollicular areas, the skin may have a “salt-and-pepper” appearance, most prominently on the scalp, upper back, and chest. Dermal sclerosis obliterating hair follicles, sweat glands, and eccrine and sebaceous glands cause hair loss, decreased sweating, and dry and itchy skin on the extremities. Transverse creases on the dorsum of the fingers disappear (**Fig. 353-7**). Fixed flexion contractures of the fingers cause reduced hand mobility and lead to muscle atrophy. Skin and subjacent tendon fibrosis accounts for fixed contractures of the wrists, elbows, and knees. Thick ridges at the neck due to firm adherence of skin to the underlying platysma muscle interfere with neck extension.

FIGURE 353-7

Sclerodactyly. Note skin induration on the fingers, and fixed flexion contractures of proximal interphalangeal joints, in a patient with limited cutaneous systemic sclerosis (lcSSc).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

In established SSc, the face assumes a characteristic “mauskopf” appearance with taut and shiny skin, loss of wrinkles, and occasionally an expressionless facies due to reduced mobility of the eyelids, cheeks, and mouth. Thinning of the lips with accentuation of the central incisor teeth and prominent perioral radial furrowing (rhytides) complete the picture. Reduced oral aperture (microstomia) interferes with eating and oral hygiene. The nose assumes a pinched, beak-like appearance. In late-stage disease, the skin becomes thin and atrophic, and is firmly bound to the subcutaneous fat (tethering). Dilated skin capillaries 2–20 mm in diameter (telangiectasiae), reminiscent of hereditary hemorrhagic telangiectasia, are frequently seen on the face, hands, lips, and oral mucosa (**Fig. 353-8**). The number of telangiectasias correlates with the severity of microvascular disease, including PAH. Breakdown of atrophic skin leads to chronic ulcerations at the extensor surfaces of the proximal interphalangeal joints, the volar pads of the fingertips, and bony prominences such as elbows and malleoli. Ulcers are often painful, heal slowly, and become secondarily infected, resulting in osteomyelitis. Healing of ischemic fingertip ulcerations leaves characteristic fixed digital “pits.” Loss of soft tissue at the fingertips due to ischemia may be associated with striking resorption of the terminal phalanges (acro-osteolysis) (**Fig. 353-9**).

FIGURE 353-8

Cutaneous vascular changes. **A.** Vascular changes at the nailfold in lcSSc. **B.** Telangiectasia on the face.



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 353-9

Acro-osteolysis. Note dissolution of distal terminal phalanges, commonly associated with ischemia, in a patient with long-standing limited cutaneous systemic sclerosis (lcSSc) and Raynaud's phenomenon.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com Copyright © McGraw-Hill Education. All rights reserved.

Dystrophic calcifications in the skin, subcutaneous, and soft tissues (calcinosis cutis) in the presence of normal serum calcium and phosphate levels occur in up to 40% of patients, most commonly in those with long-standing anti-centromere antibody-positive lcSSc. Calcific deposits, composed of calcium hydroxyapatite crystals, vary in size from tiny punctate lesions to large conglomerate masses can be readily visualized on plain radiographs, or dual-energy CT. These deposits occur when calcium precipitates in tissue damaged by inflammation, hypoxia, or local trauma. Common locations include the finger pads, palms, extensor surfaces of the forearms, and the olecranon and prepatellar bursae (**Fig. 353-10**). They can cause pain and nerve compression, ulcerate through the overlying skin with drainage of chalky white material, and secondary infections. Paraspinal sheet calcifications may cause neurologic complications.

FIGURE 353-10

Calcinosis cutis. Note soft tissue calcific deposit breaking through the skin in a patient with limited cutaneous systemic sclerosis (lcSSc).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

PULMONARY FEATURES

The two principal forms of lung involvement in SSc, ILD, and pulmonary vascular disease are frequent and account for a majority of SSc-related deaths. Survival is particularly poor in SSc patients with concurrent presence of these two processes. Less common pulmonary complications of SSc include aspiration pneumonitis complicating chronic gastroesophageal reflux, pulmonary hemorrhage due to endobronchial telangiectasia, obliterative bronchiolitis, pleural reactions, restrictive physiology due to chest wall fibrosis, spontaneous pneumothorax, and drug-induced lung toxicity. The incidence of lung cancer is increased in SSc.

Interstitial Lung Disease

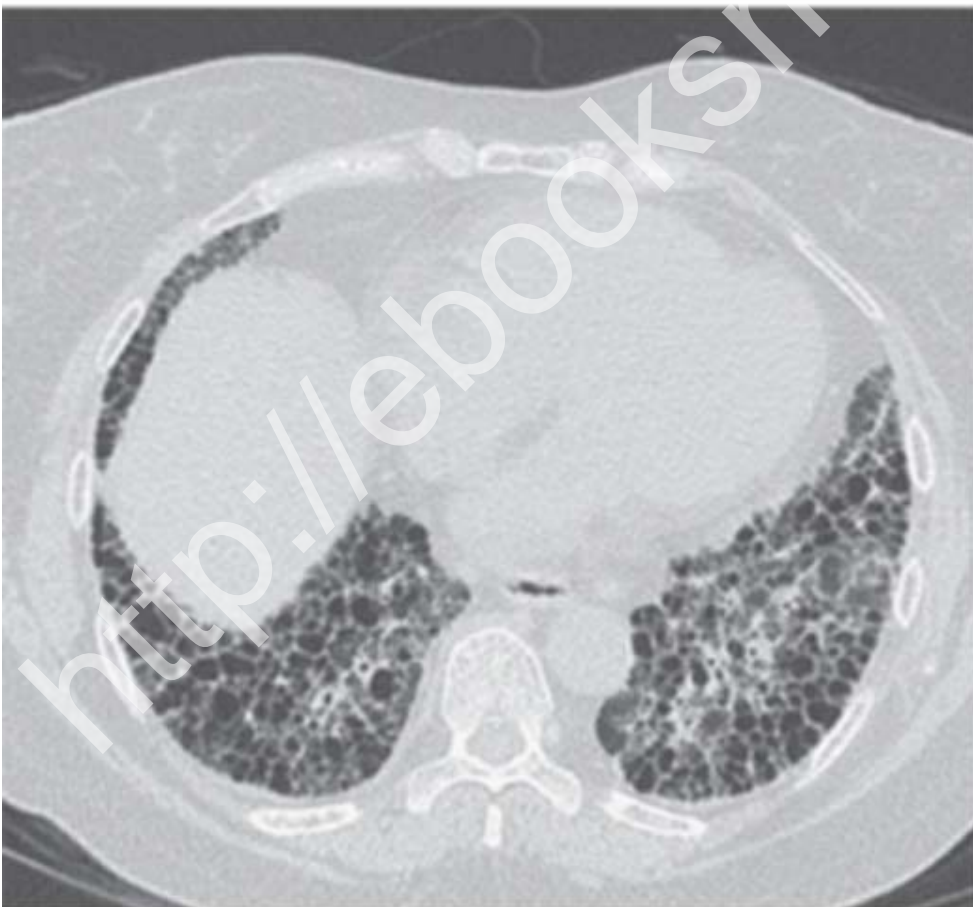
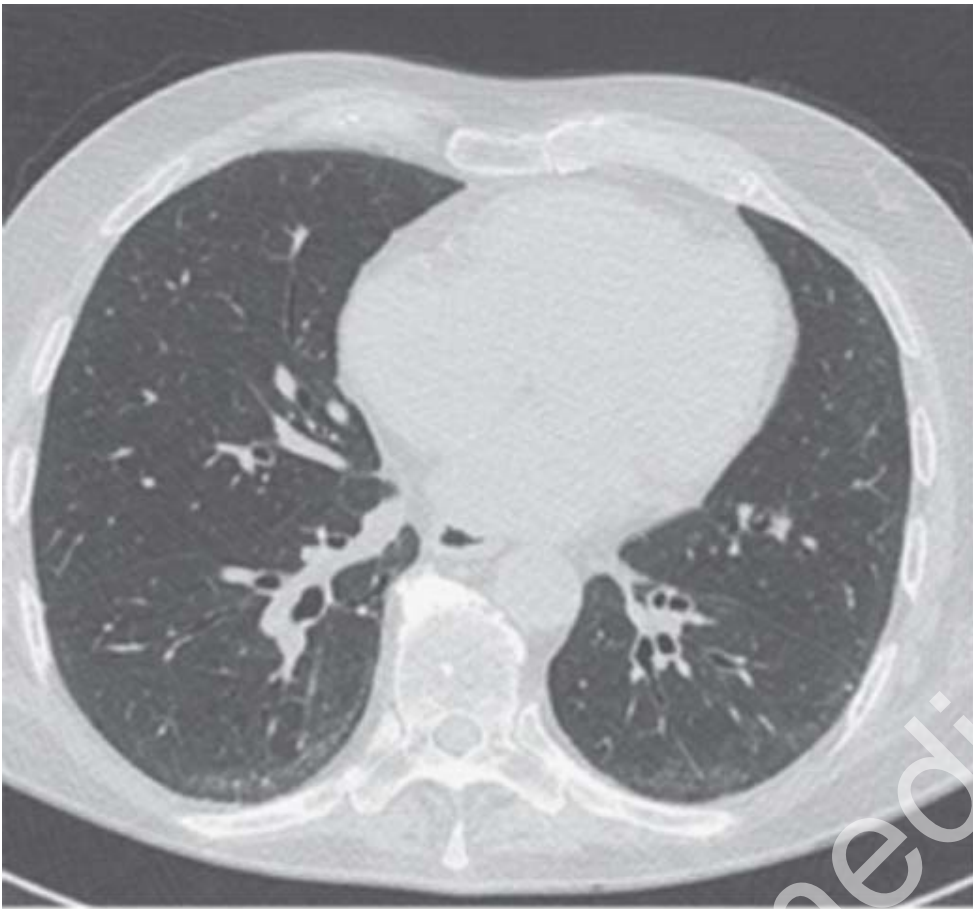
While evidence of ILD can be found in up to 65% of SSc patients by high-resolution computed tomography (HRCT), clinically significant ILD develops in 16–43%; the frequency varies depending on the detection method used. Risk factors for ILD include male sex, African-American race, diffuse skin involvement, severe gastroesophageal reflux, and the presence of topoisomerase I autoantibodies; in contrast, anti-centromere antibody-positive patients have a reduced risk of ILD. Additional risk factors include low forced vital capacity (FVC) or single-breath diffusing capacity of the lung for carbon monoxide (DLco) at initial presentation. Esophageal dilation with chronic acid reflux in SSc cause micro-aspiration, a risk factor for the development and progression of ILD. The most rapid progression in ILD generally occurs early in the disease course (within the first 3 years), when the FVC can decline by 30% per year.

Pulmonary involvement can remain asymptomatic until it is advanced. The most common presenting respiratory symptoms—exertional dyspnea, fatigue, and reduced exercise tolerance—are subtle and slowly progressive. A chronic dry cough may be present. Physical examination may reveal fine inspiratory “Velcro” crackles at the lung bases. Pulmonary function testing (PFT) is relatively sensitive for detecting early pulmonary involvement, and typically shows a restrictive ventilatory defect ($FV < 70\%$ predicted and/or FEV_1/FVC ratio > 0.8), reduced total lung capacity (TLC) and diffusing capacity (DLco). A reduction in DLco that is significantly out of proportion to the reduction in lung volumes should raise suspicion for pulmonary vascular disease, but may also be due to anemia. Oxygen desaturation with exercise is common.

Chest radiography can be used as an initial screening tool to rule out infection and other causes of pulmonary involvement; however, compared to HRCT, it is relatively insensitive for detection of early ILD. It may demonstrate lower lobe subpleural reticular linear opacities and ground-glass opacifications, even in asymptomatic patients with normal PFTs (**Fig. 353-11**). Additional HRCT findings include mediastinal lymphadenopathy, pulmonary nodules, traction bronchiectasis, and uncommonly, honeycomb changes. The extent of interstitial changes on chest HRCT is a predictor of ILD progression and mortality. Bronchoalveolar lavage (BAL) can demonstrate inflammatory cells in the lower respiratory tract, and may be useful for ruling out tuberculosis and other infections. However, BAL does not appear to be useful for SSc diagnosis or for identifying reversible alveolitis, and is used primarily for research. Lung biopsy is indicated only in patients with atypical findings on chest radiographs. The histologic pattern on lung biopsy may predict the risk of progression of ILD, with NSIP, carrying a better prognosis than UIP.

FIGURE 353-11

Chest CT in systemic sclerosis. *Top panel:* Early interstitial lung disease with subpleural reticulations and ground glass opacities in the lower lobes. Patient in supine position. ***Bottom panel:*** Extensive lung fibrosis with coarse reticular honeycombing, and traction bronchiectasis. Note dilated esophagus. (*Courtesy of Rishi Agrawal, Northwestern University.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com Copyright © McGraw-Hill Education. All rights reserved.

Pulmonary Arterial Hypertension PAH resulting from vascular remodeling of small (<500 μ m) pulmonary arteries develops in 8–12% of patients with SSc, and occurs as an isolated abnormality or in association with ILD. PAH is defined hemodynamically as a mean pulmonary artery pressure ≥ 25 mmHg with a pulmonary capillary wedge pressure ≤ 15 mmHg and pulmonary vascular resistance > 3 Wood units. The natural history of SSc-associated PAH is variable, but often follows a downhill course with onset of right heart failure. The 3-year survival of SSc patients with untreated PAH is $< 50\%$. Risk factors include lcSSc, high numbers of cutaneous telangiectasia, older age at disease onset, and the presence of antibodies to centromere, U1-RNP, U3-RNP (fibrillarin), and B23. Mutations in the BMPR2 gene associated with idiopathic PAH are not found in patients with SSc-PAH.

Although patients with PAH are often asymptomatic in early stages, they may present with nonspecific symptoms of exertional dyspnea and reduced exercise capacity. With progression, angina, near-syncope, and symptoms and signs of right-sided heart failure appear. Physical examination may show tachypnea, a loud pulmonic component of the S₂ heart sound, pulmonic/tricuspid regurgitation murmur, palpable right ventricular heave, elevated jugular venous pressure, and dependent edema. Doppler echocardiography provides a noninvasive screening method for estimating the pulmonary arterial pressure. In light of the poor prognosis of untreated PAH and better therapeutic response in patients with early diagnosis, all SSc patients should be screened for PAH at initial evaluation, followed by annual evaluation. Estimated pulmonary artery systolic pressure > 40 mmHg at rest or tricuspid regurgitation jet velocities > 3 m/sec suggest PAH. PFT may show a reduced DLco in isolation or out of proportion with the severity of restriction. Because echocardiography can over- or underestimate pulmonary artery pressures, cardiac catheterization is the gold standard required to confirm the diagnosis of suspected PAH, to assess its severity, including the degree of right heart dysfunction, to rule out veno-occlusive disease and other cardiac (post-capillary) causes of pulmonary hypertension, and to provide prognostic parameters. Yearly echocardiographic screening for PAH is recommended in most patients; an isolated decline in DLco may also be indicative of developing PAH. Distinguishing PAH from pulmonary hypertension secondary to pulmonary fibrosis and hypoxia in SSc can be difficult. Serum levels of N-terminal pro-brain natriuretic peptide (NT proBNP) correlate with the presence and severity of PAH in SSc, as well as survival. While NT proBNP measurements can be useful in screening for PAH and in monitoring the response to treatment, elevated levels are not specific for PAH and also occur in other forms of right and left heart disease. Despite more favorable hemodynamics, the prognosis of SSc-associated PAH is worse, and treatment response poorer, than that of idiopathic PAH, most likely due to frequent concurrence of ILD and cardiac complications in these patients.

GASTROINTESTINAL INVOLVEMENT

Involvement of the gastrointestinal tract, which can affect any level, occurs in up to 90% of SSc patients with both lcSSc and dcSSc disease (**Table 353-6**). The pathologic findings of fibrosis, smooth muscle atrophy, and obliterative small-vessel vasculopathy are similar throughout the length of the gastrointestinal tract, and contribute to reduced quality of life, malnutrition, and increased mortality.

TABLE 353-6

Prominent Gastrointestinal Manifestations of SSc and Their Management

SITE	PRINCIPAL MANIFESTATION	MANAGEMENT
Oropharynx	Diminished oral aperture Dry mouth Periodontitis, gingivitis swallowing	Periodontal care Artificial saliva Swallowing therapy
Esophagus	Reflux Dysphagia Strictures Barret's metaplasia	Lifestyle modifications Prokinetic drugs proton pump inhibitors Endoscopic procedures
Stomach	Gastroparesis Gastric antral vascular ectasia (GAVE, watermelon stomach)	Prokinetic agents Endoscopic laser cryotherapy
Small and large intestines	Bacterial overgrowth Diarrhea/constipation Pseudo-obstruction Pneumatosis intestinalis Malabsorption Colonic pseudodiverticula	Laxatives Prokinetic agents Rotating antibiotics Octreotide Parenteral nutritional support
Anorectum	Sphincter incompetence	Biofeedback, sacral nerve stimulation, surgery

Upper Gastrointestinal Tract Involvement

Decreased oral aperture interferes with regular dental hygiene. Teeth are loosened due to loss of periodontal ligament attaching teeth to the alveolar bone. Additional oropharyngeal manifestations due to a combination of xerostomia, shortened frenulum, and resorption of the mandibular condyles are frequent and cause much distress. Most patients have symptoms of gastroesophageal reflux disease (GERD): heartburn, regurgitation, and dysphagia. A combination of reduced lower esophageal sphincter pressure resulting in reflux, impaired esophageal clearance of refluxed gastric contents due to diminished motility, and delayed gastric emptying accounts for GERD. Calcium channel antagonists and phosphodiesterase inhibitors used to treat Raynaud's phenomenon can further aggravate reflux. Esophageal manometry shows abnormal motility in most patients, even in the absence of symptoms. Extra-esophageal manifestations of GERD include hoarseness, chronic cough, and microaspiration, which can result in infections and may aggravate underlying ILD. Chest CT characteristically shows a dilated patulous esophagus with

intraluminal air. Endoscopy may be necessary to rule out opportunistic infections with *Candida*, herpes virus, and CMV. Severe erosive esophagitis may be found on endoscopy in patients with minimal symptoms. Esophageal strictures and Barrett's esophagus may complicate chronic GERD. Because Barrett's metaplasia is associated with increased risk of adenocarcinoma, SSc patients with Barrett's require regular surveillance endoscopy with biopsy.

Gastroparesis with early satiety, abdominal distention, and aggravated reflux symptoms are common. Barium contrast studies are neither sensitive nor specific for evaluation of gastric involvement in SSc. Gastric antral vascular ectasia (GAVE) in the antrum may occur. These subepithelial lesions, reflecting the diffuse small-vessel vasculopathy of SSc, are described as "watermelon stomach" due to their endoscopic appearance. Patients with GAVE can have recurrent episodes of gastrointestinal bleeding, resulting in chronic unexplained anemia.

Lower Gastrointestinal Tract and Anorectal Involvement

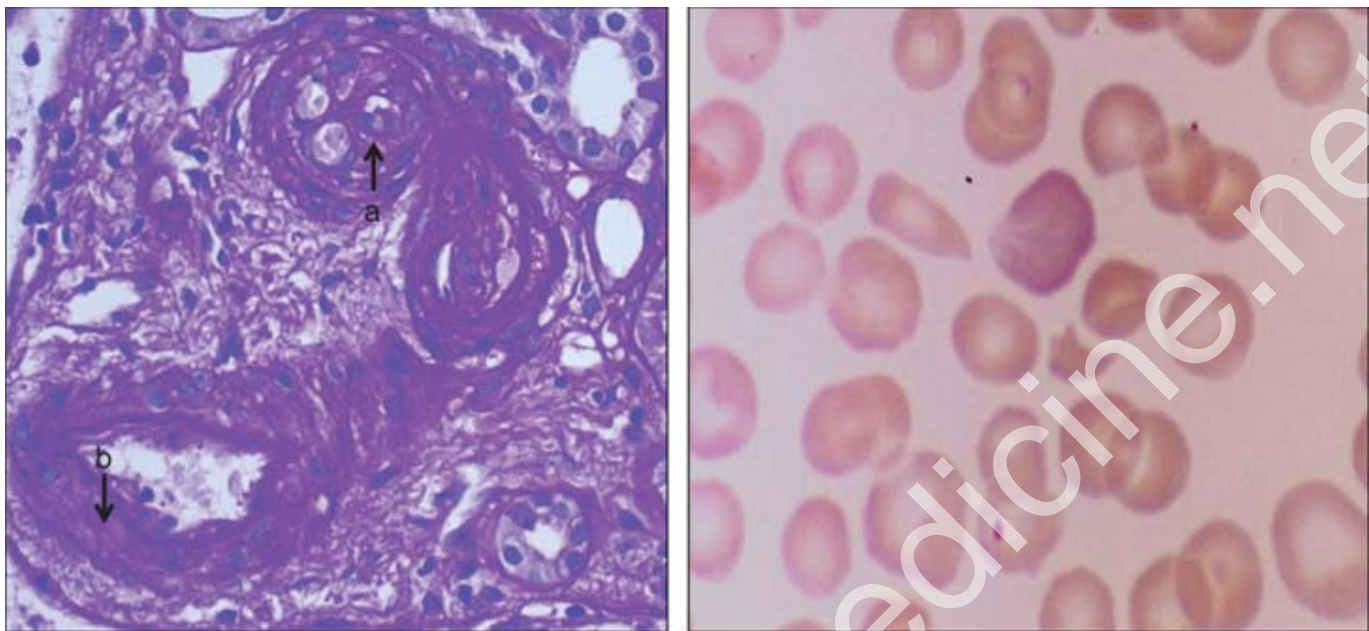
Weight loss and malnutrition due to impaired intestinal motility, malabsorption, and chronic diarrhea secondary to bacterial overgrowth are common. Fat and protein malabsorption and vitamin B₁₂ and vitamin D deficiencies ensue, and may be further exacerbated by pancreatic insufficiency. Disturbed intestinal motor function can also lead to intestinal pseudo-obstruction, with symptoms that are indistinguishable from those of delayed gastric emptying. Patients present with recurrent episodes of acute abdominal pain, nausea, and vomiting, and radiographic studies show acute intestinal obstruction. A major diagnostic challenge is differentiating pseudo-obstruction, which responds to supportive care and intravenous nutritional supplementation, from mechanical obstruction. Colonic involvement may result in severe constipation, occasionally complicated by sigmoid volvulus. Fecal incontinence, gastrointestinal bleeding from telangiectasia, and rectal prolapse, can occur. In late-stage SSc, wide-mouth sacculations or diverticula occur in the colon, occasionally causing perforation and bleeding. An occasional radiologic finding is pneumatosis cystoides intestinalis due to air trapping in the bowel wall that may rarely rupture and cause benign pneumoperitoneum. Although the liver is rarely affected, primary biliary cirrhosis may coexist with SSc.

RENAL INVOLVEMENT: SCLERODERMA RENAL CRISIS

Scleroderma renal crisis presents with accelerated hypertension accompanied by acute kidney injury and progressive failure. This acute life-threatening complication of SSc occurs in 10–15% of patients, generally within 4 years of disease onset. Rarely, scleroderma renal crisis can be the initial presenting manifestation of SSc. Prior to the advent of angiotensin-converting enzyme (ACE) inhibitors, short-term survival in scleroderma renal crisis was <10%. The pathogenesis involves obliterative vasculopathy and luminal narrowing of the renal arcuate and interlobular arteries, with consequent intravascular hemolysis, along with evidence of activation of the complement pathways (**Fig. 353-12**). Progressive reduction in renal blood flow, aggravated by vasospasm, leads to juxtaglomerular renin secretion and activation of Angiotensin II, with further renal vasoconstriction resulting in a vicious cycle that culminates in accelerated hypertension. Risk factors for scleroderma renal crisis include African-American race, male sex, and diffuse or progressive skin involvement. Up to 50% of patients with scleroderma renal crisis have anti-RNA polymerase III antibodies, whereas patients with anti-centromere antibodies appear to be protected from this complication. Palpable tendon friction rubs, pericardial effusion, new unexplained anemia, and thrombocytopenia may be harbingers of impending scleroderma renal crisis. High-risk patients with early SSc should monitor their blood pressure daily. Because glucocorticoid use is associated with scleroderma renal crisis, **prednisone** in high-risk SSc patients should be taken only when absolutely required and at low doses (<10 mg/d).

FIGURE 353-12

Renal changes in scleroderma renal crisis. **A.** Renal biopsy demonstrating intimal proliferation and myxoid changes in medium-sized renal arteries. **B.** Film demonstrating fragmentation of red blood cells due to intravascular hemolysis in scleroderma renal crisis. (Courtesy of Drs. Edward Stern and Christopher Denton, Royal Free Hospital, London, UK.)

**A****B**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com
Copyright © McGraw-Hill Education. All rights reserved.

Patients characteristically present with accelerated hypertension (generally >150/90 mmHg) and progressive oliguric renal insufficiency. However, ~10% of patients with scleroderma renal crisis present with normal blood pressure. Normotensive renal crisis is generally associated with a poor outcome. Headache, blurred vision, congestive heart failure, and pulmonary edema may accompany elevation of blood pressure. Urinalysis typically shows mild proteinuria, granular casts, and microscopic hematuria; moderate thrombocytopenia and microangiopathic hemolysis with fragmented red blood cells can be seen. Progressive oliguric renal failure over several days generally follows. Scleroderma renal crisis is occasionally misdiagnosed as thrombotic thrombocytopenic purpura (TTP) or other forms of thrombotic microangiopathy. In such cases, renal biopsy and measuring vWF-cleaving protease activity may be of some benefit. Oliguria or a creatinine >3 mg/dL at presentation predicts poor outcome (permanent hemodialysis and mortality), as do biopsy findings of vascular thrombosis and glomerular ischemic collapse. Rarely, crescentic glomerulonephritis occurs in the setting of SSc and may be associated with myeloperoxidase-specific antineutrophil cytoplasmic antibodies. Membranous glomerulonephritis may occur in patients treated with d-penicillamine. Asymptomatic renal function impairment occurs in up to half of SSc patients. Such subclinical renal involvement is associated with other vascular manifestations of SSc and rarely progresses.

CARDIAC INVOLVEMENT

Although it is often silent, variable cardiac involvement in SSc is detected in 10–50% of patients screened with sensitive diagnostic tools. Clinical cardiac involvement, more frequent in dcSSc than in lcSSc, may be primary or secondary to PAH, ILD, or renal involvement, and is associated with poor outcomes. The endocardium,

myocardium, and pericardium may each be affected separately or together. Pericardial involvement is manifested as pericarditis, pericardial effusions, constrictive pericarditis, and rarely, cardiac tamponade. Conduction system fibrosis occurs commonly and may be silent or manifested by heart block. Arrhythmias including premature ventricular contractions, atrial fibrillation, and supraventricular tachycardia are common. Microvascular involvement, recurrent vasospasm, and ischemia-reperfusion injury contribute to patchy myocardial fibrosis, resulting in asymptomatic systolic or diastolic left ventricular dysfunction that may progress to overt heart failure. Acute or subacute myocarditis leading to left ventricular dysfunction may occur, and diagnosis requires cardiac magnetic resonance imaging (MRI) or endomyocardial biopsy. While conventional echocardiography has low sensitivity for detecting preclinical heart involvement in SSc, newer modalities such as tissue Doppler echocardiography (TDE), cMRI, and nuclear imaging (single photon emission CT [SPECT]) reveal a high prevalence of abnormal myocardial function or perfusion. The serum levels of N-terminal pro-BNP, a ventricular hormone elevated in SSc-PAH, may also have utility as markers of primary cardiac involvement.

Musculoskeletal Complications

Musculoskeletal complications are very common in SSc. Carpal tunnel syndrome may be a presenting disease manifestation. Generalized arthralgia and stiffness are prominent in early disease. Mobility of both small and large joints is progressively impaired, and fixed contractures develop at the proximal interphalangeal joints and wrists. Large joint contractures, seen in patients with dcSSc, are frequently accompanied by tendon friction rubs characterized by coarse leathery crepitation heard or palpated upon passive joint movement, that are due to extensive fibrosis and adhesion of the tendon sheaths and fascial planes at the affected joint. Tendon friction rubs are associated with increased risk for renal and cardiac complications and reduced survival. Synovitis detected by ultrasound or MRI is common; occasional SSc patients develop erosive polyarthritis in the hands, and some have a seropositive rheumatoid arthritis overlap. Muscle weakness is common and multifactorial: deconditioning, disuse atrophy, malnutrition, inflammation, and fibrosis may all contribute. A chronic non-inflammatory myopathy characterized by atrophy and fibrosis with mildly elevated muscle enzymes can be seen in late-stage SSc. Bone resorption in the terminal phalanges causes loss of the distal tufts (acro-osteolysis) (Fig. 353-5). Resorption of the mandibular condyles can lead to bite difficulties. Osteolysis can also affect the ribs and distal clavicles.

LESS RECOGNIZED DISEASE MANIFESTATIONS

Dry eyes and dry mouth (sicca complex) are common in SSc. Biopsy of the minor salivary glands shows fibrosis rather than focal lymphocytic infiltration characteristic of primary Sjögren's syndrome (Chap. 354). Hypothyroidism resulting from Graves' or Hashimoto's disease is common, particularly in lcSSc, and may be under-recognized. Whereas the central nervous system is generally spared, unilateral or bilateral sensory trigeminal neuropathy can occur. Erectile dysfunction is a frequent, and occasionally initial, disease manifestation. Inability to attain or maintain penile erection is due to vascular insufficiency and fibrosis of corporeal smooth muscle. Sexual performance is also adversely affected in women. While fertility is not impaired in SSc, pregnancy is associated with higher risk of adverse fetal outcomes. Furthermore, cardiopulmonary involvement may worsen during pregnancy, and new onset of scleroderma renal crisis has been described.

Cancer

Epidemiologic studies indicate an increased cancer risk in SSc. Lung cancer and esophageal adenocarcinoma typically occur in the setting of long-standing ILD or GERD and may be caused by chronic inflammation and repair. In contrast, breast, lung, and ovarian carcinomas and lymphomas tend to occur in close temporal association with the onset of SSc, particularly in patients who have autoantibodies to RNA polymerase III. In this scenario, SSc may represent a paraneoplastic syndrome triggered by the anti-tumor immune response.

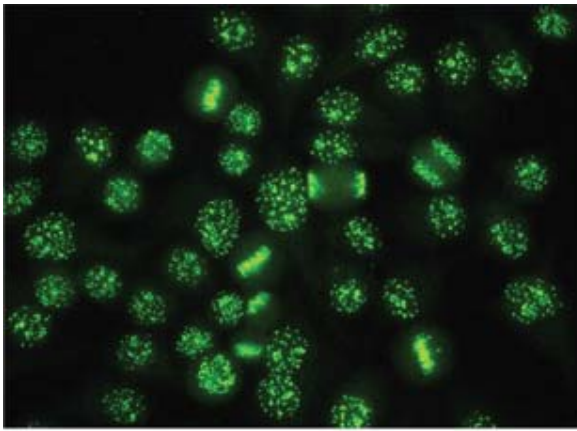
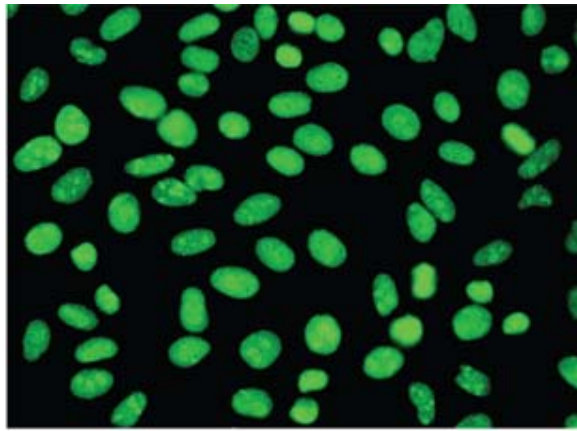
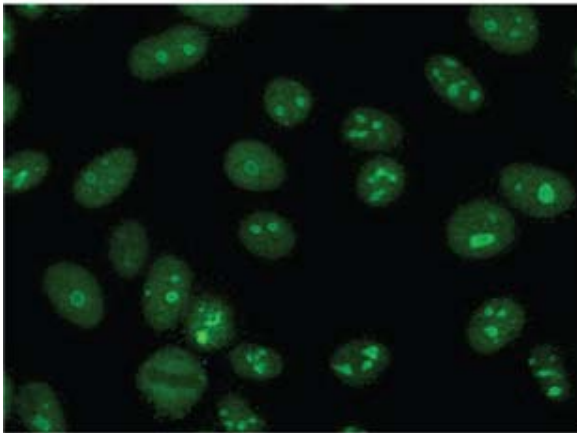
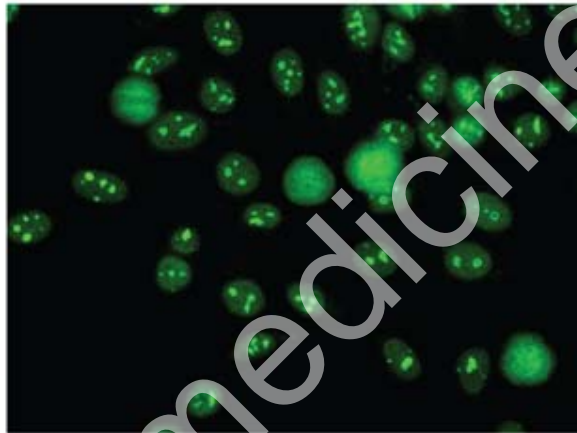
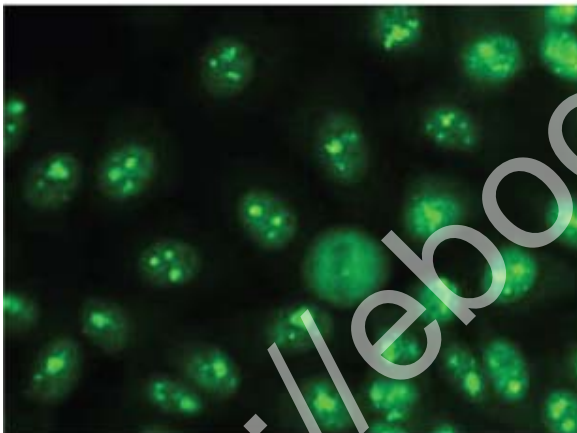
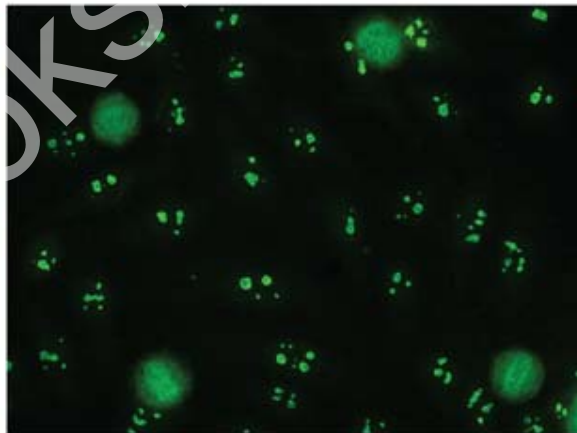
LABORATORY EVALUATION AND BIOMARKERS

Mild microcytic anemia is frequent and may indicate gastrointestinal bleeding caused by GAVE or chronic esophagitis. Macrocytic anemia may be caused by folate and vitamin B₁₂ deficiency due to small-bowel bacterial overgrowth and malabsorption or by drugs such as methotrexate. Microangiopathic hemolytic anemia caused by mechanical fragmentation of red blood cells during their passage through microvessels coated with fibrin or platelet thrombi is a hallmark of scleroderma renal crisis. The erythrocyte sedimentation rate (ESR) is generally normal; an elevation may signal coexisting myositis or malignancy.

Antinuclear autoantibodies are detected in almost all patients with SSc. Anti-topoisomerase I (Scl-70) and anti-centromere antibodies are mutually exclusive and each is highly specific for SSc. Topoisomerase I antibodies are associated with increased risk of ILD and poor outcomes. Anti-centromere antibodies are associated with PAH, but only infrequently with significant cardiac, pulmonary, or renal involvement. Nucleolar immunofluorescence pattern may indicate antibodies to U3-RNP (fibrillarin), Th/To, or PM/Scl, whereas speckled immunofluorescence indicates antibodies to RNA polymerase III (**Fig. 353-13**).

FIGURE 353-13

SSc-associated autoantibodies: immunofluorescence patterns. Indirect immunofluorescence on HEp-2 substrate shows distinct patterns: **A.** anti-centromere; **B.** anti-Scl-70/topoisomerase I; **C.** anti-PM/Scl; **D.** anti-Th/To; **E.** anti-RNA polymerase III; **F.** anti-fibrillarin/U3RNP antibodies. Except for anti-centromere (discrete dots in metaphase nucleus), variations of nucleolar staining are clues to autoantibody specificity. However, immunoassays employing purified autoantigens are recommended to confirm specificity of these autoantibodies. (*Courtesy of Marvin Fritzler and Susan Copple, Inova Diagnostics Inc., San Diego, California.*)

**A****B****C****D****E****F**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition: www.accessmedicine.com Copyright © McGraw-Hill Education. All rights reserved.

DIAGNOSIS, STAGING, AND MONITORING

The diagnosis of SSc is made primarily on clinical grounds and is generally straightforward in patients with established disease. The presence of skin induration with a characteristic symmetric distribution pattern associated with typical visceral organ manifestations establishes the diagnosis with a high degree of certainty. In lcSSc, a history of Raynaud's phenomenon and GERD symptoms, coupled with sclerodactyly and nailfold capillary changes, often in combination with cutaneous telangiectasia and calcinosis cutis, help to establish the diagnosis. Primary

Raynaud's disease is a benign condition that must be differentiated from early or limited SSc. Nailfold microscopy is particularly helpful in this situation, because in contrast to SSc, nailfold capillaries are normal. Diagnosing SSc at an early stage may be a challenge. In dcSSc, initial symptoms are often nonspecific, Raynaud's phenomenon may be absent, and physical examination may only show upper extremity edema and puffy fingers. Patients with early SSc might be diagnosed as arthritis, SLE, myositis, or, most commonly, undifferentiated connective tissue disease. Within weeks to months, Raynaud's phenomenon and advancing skin induration appear. SSc-specific autoantibodies provide a high degree of diagnostic certainty. Raynaud's phenomenon with fingertip ulcerations or other evidence of digital ischemia, coupled with telangiectasia, distal esophageal dysmotility, unexplained ILD or PAH, or accelerated hypertension with renal failure in the absence of clinically evident skin induration, suggests the diagnosis of SSc *sine* scleroderma.

APPROACH TO THE PATIENT

APPROACH TO THE PATIENT: Management of Systemic Sclerosis

OVERVIEW: GENERAL PRINCIPLES

To date, no therapy has been shown to significantly alter the natural history of SSc. In contrast, multiple interventions are highly effective in alleviating the symptoms, slowing the progression of the cumulative organ damage, and reducing disability. A significant reduction in disease-related mortality has been noted during the past 25 years. In light of the marked heterogeneity in disease manifestations, and natural history, the management of SSc mandates a "personalized medicine" approach that is specifically tailored to each individual patient's unique needs.

The following general principles should guide management ([Table 353-7](#)): prompt and accurate diagnosis; classification and risk stratification based on clinical and laboratory evaluation, including prognostic and predictive biomarkers; early recognition of organ-based complications and assessment of their extent, severity, and likelihood of deterioration; regular monitoring for disease progression, new complications, and response to therapy; adjusting therapy; and patient education. In order to minimize irreversible organ damage, management should be proactive, with regular screening and initiation of appropriate, intervention at the earliest possible opportunity. In light of the complex and multisystemic nature of the SSc, a team-oriented management approach integrating appropriate specialists should be pursued. Generally, a combination of drugs that impact different aspects of the disease is used. Patients should be encouraged to become familiar with potential complications and understand therapeutic options, including interventional trials, and natural history, and empowered to partner with their treating physicians. This requires a long-term relationship between patient and physician, with ongoing counseling, encouragement, and two-way dialogue.

DISEASE-MODIFYING THERAPY: IMMUNOSUPPRESSIVE AGENTS

Immuno-suppressive agents used in other autoimmune diseases have generally shown modest or no benefit in SSc. Glucocorticoids alleviate stiffness and aching in early inflammatory-stage dcSSc, but do not influence the progression of skin or internal organ involvement. Since their use is associated with an increased risk of scleroderma renal crisis, glucocorticoids should be given only when absolutely necessary, at the lowest dose possible, and for brief periods only.

Cyclophosphamide has been extensively studied in light of its efficacy in the treatment of vasculitis (**Chap. 356**), SLE (**Chap. 349**), and other autoimmune diseases (**Chap. 348**). Both oral and intravenous cyclophosphamide have been shown to reduce the progression of SSc-associated ILD, with stabilization and, rarely, modest improvement of pulmonary function, HRCT findings, respiratory symptoms, and skin induration. The benefits of cyclophosphamide need to be balanced against its potential toxicity, including bone marrow suppression, opportunistic infections, hemorrhagic cystitis and bladder cancer, premature ovarian failure, and late secondary malignancies.

Methotrexate had modest effect on SSc skin involvement in small studies. Mycophenolate mofetil was evaluated in both open label and randomized control trials. Both skin induration and ILD improved in patients treated with MMF, and the drug was well tolerated. Tocilizumab, a monoclonal antibody directed against the IL-6 receptor that blocks IL-6 signaling, also showed benefit in randomized SSc trials. Open-label studies and small trials provide support for the use of rituximab, a monoclonal antibody directed against the mature B cell marker CD20, along with extracorporeal photopheresis and IV immunoglobulin. Randomized trials in SSc evaluating the efficacy of **abatacept**, a fusion protein that inhibits T cell co-stimulation and function, are on-going. The use of **cyclosporine**, azathioprine, plaquenil, thalidomide, and rapamycin is currently not well supported by the literature. Intensive immune ablation using high-dose chemotherapy, followed by autologous stem cell reconstitution, in selected patients was associated with durable remission and improved long-term survival in randomized clinical trials; however, this regimen carries potential morbidity and mortality, as well as significant cost.

Antifibrotic Therapy

Because tissue fibrosis underlies organ damage in SSc, drugs that interfere with the fibrotic process represent a rational therapeutic approach. In older retrospective studies, D-penicillamine was shown to stabilize skin induration, prevent new internal organ involvement, and improve survival. However, a randomized-controlled clinical trial in early active SSc found no difference in the extent of skin involvement between patients treated with standard-dose (750 mg/d) or very low-dose (125 mg every other day) D-penicillamine. Recent clinical trials show benefit of pirfenidone and of nintedanib in patients with idiopathic pulmonary fibrosis, with significant slowing of the loss of lung function. Whether these anti-fibrotic drugs have comparable efficacy and tolerability in patients with SSc-associated ILD and other fibrotic manifestations of the disease is under investigation.

Vascular Therapy

The goal of Raynaud's therapy is to control episodes, prevent and enhance the healing of ischemic complications, and slow the progression of obliterative vasculopathy. Patients should dress warmly, minimize cold exposure, and avoid drugs that precipitate or exacerbate vasospastic episodes. Extended-release dihydropyridine calcium channel blockers such as amlodipine and diltiazem ameliorate Raynaud's phenomenon, but their use is often limited by side effects (palpitations, dependent edema, worsening gastroesophageal reflux). While ACE inhibitors do not reduce the frequency or severity of episodes, Angiotensin II receptor blockers such as losartan are effective and well tolerated. Patients with Raynaud's phenomenon unresponsive to these therapies may require the addition of α_1 -adrenergic receptor blockers (e.g., prazosin), 5-phosphodiesterase inhibitors (e.g., sildenafil), topical nitroglycerine, and intermittent IV infusions of prostaglandins. Low-dose aspirin and dipyridamole prevent platelet aggregation and may have a role as adjunctive agents. In patients with ischemic digital tip ulcerations, the endothelin-1 receptor antagonist bosentan reduces the risk of new ulcers. Digital sympathectomy and intradigital injections of botulinum type A (Botox) may be considered in patients with severe on-going ischemia. Empirical long-term therapy with statins and antioxidants may retard the progression of vascular damage and obliteration. There is limited evidence-

based information for the treatment of cardiac complications of SSc, which should be guided by specialists experienced in their diagnosis and management. While selective beta blockers such as metoprolol can precipitate vasospasm, non-dihydropyridine calcium channel blockers can be used for rate control in atrial arrhythmias, and non-selective alpha/beta blockers such as carvediol for improving myocardial perfusion and left ventricular systolic function.

TABLE 353-7

Key Principles in Management

- Establish early and accurate diagnosis.
- Detect and evaluate internal organ involvement.
- Define clinical disease stage and activity.
- Tailor individualized therapy to each patient's unique needs.
- Assess treatment response, and adjust therapy as needed; monitor for disease activity, progression and new complications.

TREATMENT

TREATMENT

TREATMENT OF SSc-ASSOCIATED ILD

ILD is a leading cause of death in patients with SSc. However, as SSc-associated ILD is not necessarily progressive, it is important to identify patients who are at high risk for disease progression in the absence of treatment. The extent of ILD on HRCT and the FVC at initial evaluation, and decline in PFTs during the preceding 12-month period, are helpful in identifying these patients. Patients at high risk for ILD should be monitored by performing PFTs every 6 months; serial HRCT imaging is not recommended. Cyclophosphamide, given IV or orally for 6 to 12 months, and mycophenolate mofetil slow the decline in lung function and improve respiratory symptoms; however, cyclophosphamide is associated with more frequent side effects. The safety and efficacy of anti-fibrotic drugs recently approved for idiopathic pulmonary fibrosis in the treatment of SSc-associated ILD are currently under investigation. In certain patients who show continued progression of ILD despite medical therapy, lung transplantation might be considered as a life-prolonging procedure, although significant GERD is a concern in SSc. Recurrence of SSc-ILD in transplanted lung allografts has not been reported.

TREATMENT OF GASTROINTESTINAL COMPLICATIONS

Because oral problems including decreased oral aperture, decreased saliva production, gum recession, periodontal disease, and teeth loss are common, regular dental care is recommended. Gastroesophageal reflux is very common and may occur in the absence of symptoms. Patients should be instructed to elevate the head of the bed, eat frequent small meals, and avoid alcohol, caffeine, and known reflux exacerbants, or meals before bedtime. Proton pump inhibitors reduce acid reflux and in patients with SSc may need to be given in relatively high doses. Prokinetic agents such as [metoclopramide](#), [erythromycin](#) (a motilin agonist), and domperidone may occasionally be helpful, but are frequently associated with side effects. Botulinum toxin injection sometimes ameliorates impaired gastric

emptying. Anti-reflux procedures such as Nissen fundoplication can result in secondary achalasia and generally should be avoided. Episodic bleeding from GAVE (watermelon stomach) may be amenable to treatment with endoscopic ablation using laser or argon plasma photocoagulation, although bleeding frequently recurs. Some patients may require enteral feeding and/or decompression via percutaneous gastrostomy or jejunostomy. Small bowel bacterial overgrowth secondary to dysmotility causes abdominal bloating and diarrhea, and may lead to malabsorption and severe malnutrition. Treatment with short courses of rotating broad-spectrum antibiotics such as [metronidazole](#), [erythromycin](#), and rifaximin can eradicate bacterial overgrowth. Small bowel hypomotility may respond to octreotide; however, pseudo-obstruction is difficult to treat. Fecal incontinence, a frequent and under-reported complication, may respond to anti-diarrheal medication, biofeedback therapy, sphincter augmentation, and sacral neuromodulation. Potential malnutrition should be routinely assessed.

TREATMENT OF PAH

In SSc, PAH carries an extremely poor prognosis and accounts for 30% of deaths. Because PAH is asymptomatic until advanced, patients with SSc should be screened at initial evaluation, and regularly thereafter. Treatment is generally started with an oral endothelin-1 receptor antagonist such as bosentan or a phosphodiesterase 5 inhibitor such as sildenafil. Recently, the soluble guanylate cyclase stimulator riociguat, which acts by increasing the production of nitric oxide, and the selective IP prostacyclin receptor agonist selexipag, were shown to improve PAH symptoms and survival. Patients may also require diuretics and digoxin. If hypoxemia is documented, supplemental oxygen should be prescribed in order to avoid secondary pulmonary vasoconstriction. Prostacyclin analogues such as epoprostenol or treprostinil can be given by continuous IV or SC infusion, or via intermittent nebulized inhalations. Combination therapy with different classes of agents acting additively or synergistically is often necessary. Lung transplantation remains an option for selected SSc patients with PAH who fail medical therapy, and 2-year survival rates (64%) are comparable to those of idiopathic ILD or PAH.

MANAGEMENT OF RENAL CRISIS

Scleroderma renal crisis is a medical emergency. Since the outcome is largely determined by the extent of renal damage at the time that aggressive therapy is initiated, prompt recognition of impending or early scleroderma renal crisis is essential, and efforts should be made to avoid its occurrence. High-risk SSc patients with early disease, extensive and progressive skin involvement, tendon friction rubs, and anti-RNA polymerase III antibodies should be instructed to monitor their blood pressure daily and report significant alterations immediately. Potentially nephrotoxic drugs should be avoided, and glucocorticoids should be used only when absolutely necessary and at low doses. Patients presenting with scleroderma renal crisis should be immediately hospitalized. Once other causes of renal disease are excluded, treatment should be started promptly with titration of short-acting ACE inhibitors, with the goal of achieving rapid normalization of the blood pressure. In patients with persistent hypertension, addition of angiotensin II receptor blockers, calcium channel blockers, endothelin-1 receptor blockers, prostacyclins, and direct renin inhibitors should be considered. Up to two-thirds of patients with scleroderma renal crisis will require dialysis. Substantial renal recovery can occur, and dialysis can be discontinued in 30–50% of the patients. Kidney transplantation is appropriate for patients unable to discontinue dialysis after 2 years. Survival of transplanted SSc patients is comparable to that of other diseases, and recurrence of renal crisis is rare.

SKIN CARE

Because skin involvement in SSc is never life-threatening and it stabilizes and may even regress spontaneously, disease management should not be dictated by its cutaneous manifestations. The inflammatory symptoms of early skin involvement can be controlled with antihistamines and short-term use of low-dose glucocorticoids (<5 mg/d of

[prednisone](#)). Cyclophosphamide and methotrexate have modest effects on skin induration. Because the skin is dry, the use of hydrophilic ointments and bath oils is encouraged, and regular skin massage is helpful. Telangiectasia, which presents a cosmetic problem, especially on the face, can be treated with pulsed dye laser. Ischemic digital ulcerations should be protected by occlusive dressing to promote healing and prevent infection. Infected skin ulcers are treated with topical antibiotics and surgical debridement. While no therapy has been shown to be effective in preventing soft tissue calcific deposits or promoting their dissolution, reports support the use of diltiazem, minocycline, bisphosphonates, and topical or IV [sodium thiosulfate](#) (STS). Other therapies that have been used for calcinosis include carbon dioxide laser, extracorporeal shock-wave lithotripsy, and surgical high-speed microdrilling.

TREATMENT OF MUSCULOSKELETAL COMPLICATIONS

Arthralgia and joint stiffness are very common and distressing manifestations in early-stage disease. Short courses of nonsteroidal anti-inflammatory agents, methotrexate, and cautious use of low-dose glucocorticoids alleviate symptoms. Physical and occupational therapy can be effective for preventing loss of musculoskeletal function and joint contractures, and should be initiated early.

COURSE

The natural history of SSc is highly variable and difficult to predict, especially in early stages of the disease. Patients with dcSSc tend to have a more rapidly progressive course and worse prognosis than those with lcSSc. Inflammatory symptoms of early dcSSc, such as fatigue, edema, joint pain and pruritus subside, and skin thickening reach a plateau at 2–4 years after disease onset. It is during the early edematous/inflammatory stage that life-threatening visceral organ involvement may develop. While existing visceral organ involvement, such as ILD, may progress even after skin involvement peaks, new organ involvement is rare. Scleroderma renal crisis generally occurs within the first 4 years of disease. In late-stage disease (>6 years), the skin is usually soft and atrophic. Skin regression characteristically occurs in an order that is the reverse of initial involvement, with softening on the trunks followed by proximal and finally distal extremities; however, sclerodactyly and fixed finger contractures generally persist. Relapse or recurrence of skin thickening after peak skin involvement has been reached is uncommon. Patients with lcSSc follow a clinical course that is markedly different than that of dcSSc. Raynaud's phenomenon typically precedes other disease manifestations by years or even decades. Visceral organ complications such as PAH generally develop late and progress slowly.

PROGNOSIS

SSc confers a substantial increase in the risk of premature death. Age- and gender-adjusted mortality rates are fivefold to eightfold higher compared to the general population, and more than half of all patients with SSc die from their disease. In one population-based study of SSc, the median survival was 11 years. In patients with dcSSc, 5- and 10-year survival rates are 70% and 55%, respectively, whereas in patients with lcSSc, 5- and 10-year survival rates are 90% and 75%, respectively. The prognosis correlates with the extent of skin involvement, which itself is a surrogate for visceral organ involvement. Major causes of death are PAH, pulmonary fibrosis, gastrointestinal involvement, and cardiac disease. Scleroderma renal crisis is associated with a 30% 3-year mortality. Lung cancer and excess cardiovascular deaths also contribute to increased mortality. Markers of poor prognosis include male gender, African-American race, older age at disease onset, extensive skin thickening with truncal involvement,

palpable tendon friction rubs, and evidence of significant or progressive visceral organ involvement. Laboratory predictors of increased mortality at initial evaluation include an elevated ESR, anemia, proteinuria, and anti-topoisomerase I antibodies. In one study, SSc patients with extensive skin involvement, vital capacity <55% predicted, significant gastrointestinal involvement (pseudo-obstruction or malabsorption), clinical evidence of cardiac involvement, or scleroderma renal crisis had a 9-year survival of <40%. The severity of PAH predicts mortality, and patients with mean pulmonary arterial pressure ≥ 45 mmHg had a 33% 3-year survival. The advent of ACE inhibitors in scleroderma renal crisis had a dramatic impact on survival, increasing from <10% at 1 year in the pre-ACE inhibitor era to >70% 3-year survival at the present time. Moreover, 10-year survival in SSc has improved from <60% in the 1970s to >66–78% in the 1990s, a trend that reflects both earlier detection and better management of complications.

LOCALIZED SCLERODERMA

The term *scleroderma* describes a group of localized skin disorders (Table 353-1). These occur more commonly in children than in adults, and in marked contrast to SSc, are generally not complicated by Raynaud's phenomenon or significant internal organ involvement. Morphea presents as solitary or multiple circular patches of thick skin or, rarely, as widespread induration (generalized or pansclerotic morphea); the fingers are generally spared. Linear scleroderma may affect subcutaneous tissues, leading to fibrosis and atrophy of supporting structures, tendons, muscle, and even bone. In children, the growth of affected long bones can be retarded. When linear scleroderma crosses large joints, significant contractures can develop.

MIXED CONNECTIVE TISSUE DISEASE

Patients who have lcSSc coexisting with features of SLE, polymyositis, and rheumatoid arthritis may have mixed connective tissue disease (MCTD). This overlap syndrome is generally associated with the presence of high titers of autoantibodies to U1-RNP. The characteristic initial presentation is Raynaud's phenomenon associated with puffy fingers and myalgia. Over time, sclerodactyly, soft tissue calcinosis, and cutaneous telangiectasia may appear. Skin rash suggestive of SLE (malar erythema, photosensitivity) or dermatomyositis (heliotrope rash on the eyelids, erythematous rash on knuckles) occur. Arthralgia is common, and some patients develop erosive polyarthritis. Pulmonary fibrosis and isolated or secondary PAH may develop. Other manifestations include esophageal dysmotility, pericarditis, Sjögren's syndrome, and renal disease, especially membranous glomerulonephritis. Laboratory evaluation shows elevated ESR and hypergammaglobulinemia. While anti-U1RNP antibodies are detected in high titers, SSc-specific autoantibodies are absent. In contrast to SSc, MCTD often responds to glucocorticoids, and the long-term prognosis is better than that of SSc. Whether MCTD is truly a distinct entity or is a subset of SLE or SSc, remains controversial.

EOSINOPHILIC FASCIITIS (DIFFUSE FASCIITIS WITH EOSINOPHILIA)

Eosinophilic fasciitis is a rare idiopathic disorder of adults associated with abrupt skin induration. The skin characteristically shows a coarse cobblestone "peau d'orange" appearance. In contrast to SSc, Raynaud's phenomenon and SSc-associated internal organ involvement and autoantibodies are absent. Furthermore, skin involvement spares the fingers. Full-thickness biopsy of the lesional skin reveals fibrosis of the subcutaneous fascia, with variable inflammation and eosinophil infiltration. In the acute phase of the illness, peripheral blood

eosinophilia may be prominent. MRI appears to be a sensitive tool for the diagnosis of eosinophilic fasciitis. Eosinophilic fasciitis can occur in association with, or preceding, various myelodysplastic syndromes or multiple myeloma. Although glucocorticoids cause prompt resolution of eosinophilia, the skin shows slow and variable improvement. The prognosis of patients with eosinophilic fasciitis who do not develop hematological complications is generally good.

FURTHER READING

Allanore Y et al: Systemic sclerosis. Nat Rev Dis Primers 1:15002, 2015.

Herzog EL et al: Interstitial lung disease associated with systemic sclerosis and idiopathic pulmonary fibrosis: How similar or distinct? Arthritis Rheum 66:1967, 2014.

Joseph CG et al: Association of the autoimmune disease scleroderma with an immunologic response to cancer. Science 343:152, 2014. [[PubMed: 24310608](#)]

Martyanov V, Whitfield ML: Molecular stratification and precision medicine in systemic sclerosis from genomic and proteomic data. Curr Opin Rheumatol 28:83, 2016. [[PubMed: 26555452](#)]

Tashkin DP et al: Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): A randomised controlled, double-blind, parallel group trial. Lancet Respir Med 4:708, 2016. [[PubMed: 27469583](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 369: Approach to the Patient with Endocrine Disorders

J. Larry Jameson

INTRODUCTION

The management of endocrine disorders requires a broad understanding of intermediary metabolism, reproductive physiology, bone metabolism, and growth. Accordingly, the practice of endocrinology is intimately linked to a conceptual framework for understanding hormone secretion, hormone action, and principles of feedback control ([Chap. 370](#)). The endocrine system is evaluated primarily by measuring hormone concentrations, arming the clinician with valuable diagnostic information. Most disorders of the endocrine system are amenable to effective treatment once the correct diagnosis is established. Endocrine deficiency disorders are treated with physiologic hormone replacement; hormone excess conditions, which usually are caused by benign glandular adenomas, are managed by removing tumors surgically or reducing hormone levels medically.

SCOPE OF ENDOCRINOLOGY

The specialty of endocrinology encompasses the study of glands and the hormones they produce. The term *endocrine* was coined by Starling to contrast the actions of hormones secreted internally (*endocrine*) with those secreted externally (*exocrine*) or into a lumen, such as the gastrointestinal tract. The term *hormone*, derived from a Greek phrase meaning “to set in motion,” aptly describes the dynamic actions of hormones as they elicit cellular responses and regulate physiologic processes through feedback mechanisms.

Unlike many other specialties in medicine, it is not possible to define endocrinology strictly along anatomic lines. The classic endocrine glands—pituitary, thyroid, parathyroid, pancreatic islets, adrenals, and gonads—communicate broadly with other organs through the nervous system, hormones, cytokines, and growth factors. In addition to its traditional synaptic functions, the brain produces a vast array of peptide hormones, and this has led to the discipline of neuroendocrinology. Through the production of hypothalamic releasing factors, the central nervous system (CNS) exerts a major regulatory influence over pituitary hormone secretion ([Chap. 371](#)). The peripheral nervous system stimulates the adrenal medulla. The immune and endocrine systems are also intimately intertwined. The adrenal hormone cortisol is a powerful immunosuppressant. Cytokines and interleukins (ILs) have profound effects on the functions of the pituitary, adrenal, thyroid, and gonads. Common endocrine diseases such as autoimmune thyroid disease and type 1

diabetes mellitus are caused by dysregulation of immune surveillance and tolerance. Less common diseases such as polyglandular failure, Addison's disease, and lymphocytic hypophysitis also have an immunologic basis.

The interdigitation of endocrinology with physiologic processes in other specialties sometimes blurs the role of hormones. For example, hormones play an important role in maintenance of blood pressure, intravascular volume, and peripheral resistance in the cardiovascular system. Vasoactive substances such as catecholamines, angiotensin II, endothelin, and nitric oxide are involved in dynamic changes of vascular tone in addition to their multiple roles in other tissues. The heart is the principal source of atrial natriuretic peptide, which acts in classic endocrine fashion to induce natriuresis at a distant target organ (the kidney). Erythropoietin, a traditional circulating hormone, is made in the kidney and stimulates erythropoiesis in bone marrow (**Chap. 59**). The kidney is also integrally involved in the renin-angiotensin axis (**Chap. 379**) and is a primary target of several hormones, including parathyroid hormone (PTH), mineralocorticoids, and vasopressin. The gastrointestinal tract produces a vast array of peptide hormones, such as cholecystokinin, ghrelin, gastrin, secretin, and vasoactive intestinal peptide, among many others. Carcinoid and islet tumors can secrete excessive amounts of these hormones, leading to specific clinical syndromes (**Chap. 80**). Many of these gastrointestinal hormones are also produced in the CNS, where their functions are poorly understood. Adipose tissue produces leptin, which acts centrally to control appetite, along with adiponectin, resistin, and other hormones that regulate metabolism. As hormones such as inhibin, ghrelin, and leptin are discovered, they become integrated into the science and practice of medicine on the basis of their functional roles rather than their tissues of origin.

Characterization of hormone receptors frequently reveals unexpected relationships to factors in nonendocrine disciplines. The growth hormone (GH) and leptin receptors, for example, are members of the cytokine receptor family. The G protein-coupled receptors (GPCRs), which mediate the actions of many peptide hormones, are used in numerous physiologic processes, including vision, smell, and neurotransmission.

PATHOLOGIC MECHANISMS OF ENDOCRINE DISEASE

Endocrine diseases can be divided into three major types of conditions: (1) hormone excess, (2) hormone deficiency, and (3) hormone resistance (**Table 369-1**).

TABLE 369-1

Causes of Endocrine Dysfunction

TYPE OF ENDOCRINE DISORDER	EXAMPLES
Hyperfunction	
Neoplastic	
Benign	Pituitary adenomas, hyperparathyroidism, autonomous thyroid or adrenal nodules, pheochromocytoma
Malignant	
Ectopic	Adrenal cancer, medullary thyroid cancer, carcinoid
Multiple endocrine neoplasia (MEN)	Ectopic ACTH, SIADH secretion MEN1, MEN2
Autoimmune	Graves' disease
Iatrogenic	Cushing's syndrome, hypoglycemia
Infectious/inflammatory	Subacute thyroiditis
Activating receptor mutations	LH, TSH, Ca^{2+} , PTH receptors, $\text{G}_s\alpha$
Hypofunction	
Autoimmune	Hashimoto's thyroiditis, type 1 diabetes mellitus, Addison's disease, polyglandular failure
Iatrogenic	
Infectious/inflammatory	Radiation-induced hypopituitarism, hypothyroidism, surgical
Hormone mutations	Adrenal insufficiency, hypothalamic sarcoidosis
Enzyme defects	GH, $\text{LH}\beta$, $\text{FSH}\beta$, vasopressin
Developmental defects	21-Hydroxylase deficiency
Nutritional/vitamin deficiency	Kallmann's syndrome, Turner's syndrome, transcription factors Vitamin D deficiency, iodine deficiency
Hemorrhage/infarction	Sheehan's syndrome, adrenal insufficiency
Hormone Resistance	
Receptor mutations	
Membrane	GH, vasopressin, LH, FSH, ACTH, GnRH, GHRH, PTH, leptin, Ca^{2+}
Nuclear	AR, TR, VDR, ER, GR, PPAR γ
Signaling pathway mutations	Albright's hereditary osteodystrophy
Postreceptor	Type 2 diabetes mellitus, leptin resistance

Abbreviations: ACTH, adrenocorticotrophic hormone; AR, androgen receptor; ER, estrogen receptor; FSH, follicle-stimulating hormone; GHRH, growth hormone–releasing hormone; GnRH, gonadotropin-releasing hormone; GR, glucocorticoid receptor; LH, luteinizing hormone; PPAR, peroxisome proliferator activated receptor; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone; TR, thyroid hormone receptor; TSH, thyroid-stimulating hormone; VDR, vitamin D receptor.

CAUSES OF HORMONE EXCESS

Syndromes of hormone excess can be caused by neoplastic growth of endocrine cells, autoimmune disorders, and excess hormone administration. Benign endocrine tumors, including parathyroid, pituitary, and adrenal adenomas, often retain the capacity to produce hormones, reflecting the fact that these tumors are relatively well differentiated. Many endocrine tumors exhibit subtle defects in their “set points” for feedback regulation. For example, in Cushing’s disease, impaired feedback inhibition of adrenocorticotrophic hormone (ACTH) secretion is associated with autonomous function. However, the tumor cells are not completely resistant to feedback, as evidenced by ACTH suppression by higher doses of [dexamethasone](#) (e.g., high-dose [dexamethasone](#) test) ([Chap. 379](#)). Similar set point defects are also typical of parathyroid adenomas and autonomously functioning thyroid nodules.

The molecular basis of some endocrine tumors, such as the multiple endocrine neoplasia (MEN) syndromes (MEN1, 2A, 2B), has provided important insights into tumorigenesis ([Chap. 381](#)). MEN1 is characterized primarily by the triad of parathyroid, pancreatic islet, and pituitary tumors. MEN2 predisposes to medullary thyroid carcinoma, pheochromocytoma, and hyperparathyroidism. The *MEN1* gene, located on chromosome 11q13, encodes a putative tumor-suppressor gene, *menin*. Analogous to the paradigm first described for retinoblastoma, the affected individual inherits a mutant copy of the *MEN1* gene, and tumorigenesis ensues after a somatic “second hit” leads to loss of function of the normal *MEN1* gene (through deletion or point mutations).

In contrast to inactivation of a tumor-suppressor gene, as occurs in MEN1 and most other inherited cancer syndromes, MEN2 is caused by activating mutations in a single allele. In this case, activating mutations of the *RET* protooncogene, which encodes a receptor tyrosine kinase, leads to thyroid C cell hyperplasia in childhood before the development of medullary thyroid carcinoma. Elucidation of this pathogenic mechanism has allowed early genetic screening for *RET* mutations in individuals at risk for MEN2, permitting identification of those who may benefit from prophylactic thyroidectomy and biochemical screening for pheochromocytoma and hyperparathyroidism.

Mutations that activate hormone receptor signaling have been identified in several GPCRs. For example, activating mutations of the luteinizing hormone (LH) receptor cause a dominantly transmitted form of male-limited precocious puberty, reflecting premature stimulation of testosterone synthesis in Leydig cells ([Chap. 384](#)). Activating mutations in these GPCRs are located predominantly in the transmembrane domains and induce receptor coupling to G_sα even in the absence of hormone. Consequently, adenylate cyclase is

activated, and cyclic adenosine monophosphate (AMP) levels increase in a manner that mimics hormone action. A similar phenomenon results from activating mutations in $G_s\alpha$. When these mutations occur early in development, they cause McCune-Albright syndrome. When they occur only in somatotropes, the activating $G_s\alpha$ mutations cause GH-secreting tumors and acromegaly ([Chap. 373](#)).

In autoimmune Graves' disease, antibody interactions with the thyroid-stimulating hormone (TSH) receptor mimic TSH action, leading to hormone overproduction ([Chap. 375](#)). Analogous to the effects of activating mutations of the TSH receptor, these stimulating autoantibodies induce conformational changes that release the receptor from a constrained state, thereby triggering receptor coupling to G proteins.

CAUSES OF HORMONE DEFICIENCY

Most examples of hormone deficiency states can be attributed to glandular destruction caused by autoimmunity, surgery, infection, inflammation, infarction, hemorrhage, or tumor infiltration ([Table 369-1](#)). Autoimmune damage to the thyroid gland (Hashimoto's thyroiditis) and pancreatic islet β cells (type 1 diabetes mellitus) is a prevalent cause of endocrine disease. Mutations in a number of hormones, hormone receptors, transcription factors, enzymes, and channels can also lead to hormone deficiencies.

HORMONE RESISTANCE

Most severe hormone resistance syndromes are due to inherited defects in membrane receptors, nuclear receptors, or the pathways that transduce receptor signals. These disorders are characterized by defective hormone action despite the presence of increased hormone levels. In complete androgen resistance, for example, mutations in the androgen receptor result in a female phenotypic appearance in genetic (XY) males, even though LH and testosterone levels are increased ([Chap. 381](#)). In addition to these relatively rare genetic disorders, more common acquired forms of functional hormone resistance include insulin resistance in type 2 diabetes mellitus, leptin resistance in obesity, and GH resistance in catabolic states. The pathogenesis of functional resistance involves receptor downregulation and postreceptor desensitization of signaling pathways; functional forms of resistance are generally reversible.

CLINICAL EVALUATION OF ENDOCRINE DISORDERS

Because most glands are relatively inaccessible, the physical examination usually focuses on the manifestations of hormone excess or deficiency as well as direct examination of palpable glands, such as the thyroid and gonads. For these reasons, it is important to evaluate patients in the context of their presenting symptoms, review of systems, family and social history, and exposure to medications that may affect the endocrine system. Astute clinical skills are required to detect subtle symptoms and signs suggestive of underlying endocrine disease. For example, a patient with Cushing's syndrome may manifest specific findings, such as central fat redistribution, skin striae, and proximal muscle weakness, in addition to features seen commonly in the general population, such as obesity, plethora, hypertension, and glucose intolerance. Similarly, the insidious onset of hypothyroidism—with mental slowing, fatigue, dry skin, and other features—

can be difficult to distinguish from similar, nonspecific findings in the general population. Clinical judgment that is based on knowledge of disease prevalence and pathophysiology is required to decide when to embark on more extensive evaluation of these disorders. Laboratory testing plays an essential role in endocrinology by allowing quantitative assessment of hormone levels and dynamics. Radiologic imaging tests such as computed tomography (CT) scan, magnetic resonance imaging (MRI), thyroid scan, and ultrasound are also used for the diagnosis of endocrine disorders. However, these tests generally are employed only after a hormonal abnormality has been established by biochemical testing.

HORMONE MEASUREMENTS AND ENDOCRINE TESTING

Immunoassays are the most important diagnostic tool in endocrinology, as they allow sensitive, specific, and quantitative determination of steady-state and dynamic changes in hormone concentrations. Immunoassays use antibodies to detect specific hormones. For many peptide hormones, these measurements are now configured to use two different antibodies to increase binding affinity and specificity. There are many variations of these assays; a common format involves using one antibody to capture the antigen (hormone) onto an immobilized surface and a second antibody, coupled to a chemiluminescent (immunochemiluminescent assay [ICMA]) or radioactive (immunoradiometric assay [IRMA]) signal, to detect the antigen. These assays are sensitive enough to detect plasma hormone concentrations in the picomolar to nanomolar range, and they can readily distinguish structurally related proteins, such as PTH from PTH-related peptide (PTHrP). A variety of other techniques are used to measure specific hormones, including mass spectroscopy, various forms of chromatography, and enzymatic methods; bioassays are now used rarely.

Most hormone measurements are based on plasma or serum samples. However, urinary hormone determinations remain useful for the evaluation of some conditions. Urinary collections over 24 h provide an integrated assessment of the production of a hormone or metabolite, many of which vary during the day. It is important to ensure complete collections of 24-h urine samples; simultaneous measurement of creatinine provides an internal control for the adequacy of collection and can be used to normalize some hormone measurements. A 24-h urine-free cortisol measurement largely reflects the amount of unbound cortisol, thus providing a reasonable index of biologically available hormone. Other commonly used urine determinations include 17-hydroxycorticosteroids, 17-ketosteroids, vanillylmandelic acid, metanephrine, catecholamines, 5-hydroxyindoleacetic acid, and calcium.

The value of quantitative hormone measurements lies in their correct interpretation in a clinical context. The normal range for most hormones is relatively broad, often varying by a factor of two- to tenfold. The normal ranges for many hormones are sex- and age-specific. Thus, using the correct normative database is an essential part of interpreting hormone tests. The pulsatile nature of hormones and factors that can affect their secretion, such as sleep, meals, and medications, must also be considered. Cortisol values increase fivefold between midnight and dawn; reproductive hormone levels vary dramatically during the female menstrual cycle.

For many endocrine systems, much information can be gained from basal hormone testing, particularly when different components of an endocrine axis are assessed simultaneously. For example, low testosterone and elevated LH levels suggest a primary gonadal problem, whereas a hypothalamic-pituitary disorder is likely if both LH and testosterone are low. Because TSH is a sensitive indicator of thyroid function, it is generally recommended as a first-line test for thyroid disorders. An elevated TSH level is almost always the result of primary hypothyroidism, whereas a low TSH is most often caused by thyrotoxicosis. These predictions can be confirmed by determining the free thyroxine level. In the less common circumstance when free thyroxine and TSH are both low, it is important to consider secondary hypopituitarism caused by hypothalamic-pituitary disease. Elevated calcium and PTH levels suggest hyperparathyroidism, whereas PTH is suppressed in hypercalcemia caused by malignancy or granulomatous diseases. A suppressed ACTH in the setting of hypercortisolemia, or increased urine-free cortisol, is seen with hyperfunctioning adrenal adenomas.

It is not uncommon, however, for baseline hormone levels associated with pathologic endocrine conditions to overlap with the normal range. In this circumstance, dynamic testing is useful to separate the two groups further. There are a multitude of dynamic endocrine tests, but all are based on principles of feedback regulation, and most responses can be rationalized based on principles that govern the regulation of endocrine axes. *Suppression tests* are used in the setting of suspected endocrine hyperfunction. An example is the [dexamethasone](#) suppression test used to evaluate Cushing's syndrome ([Chaps. 373 and 379](#)). *Stimulation tests* generally are used to assess endocrine hypofunction. The ACTH stimulation test, for example, is used to assess the adrenal gland response in patients with suspected adrenal insufficiency. Other stimulation tests use hypothalamic-releasing factors such as corticotropin-releasing hormone (CRH) and growth hormone-releasing hormone (GHRH) to evaluate pituitary hormone reserve ([Chap. 373](#)). Insulin-induced hypoglycemia evokes pituitary ACTH and GH responses. Stimulation tests based on reduction or inhibition of endogenous hormones are now used infrequently. Examples include [metyrapone](#) inhibition of cortisol synthesis and clomiphene inhibition of estrogen feedback.

SCREENING AND ASSESSMENT OF COMMON ENDOCRINE DISORDERS

Many endocrine disorders are prevalent in the adult population ([Table 369-2](#)) and can be diagnosed and managed by general internists, family practitioners, or other primary health care providers. The high prevalence and clinical impact of certain endocrine diseases justifies vigilance for features of these disorders during routine physical examinations; laboratory screening is indicated in selected high-risk populations.

TABLE 369-2

Examples of Prevalent Endocrine and Metabolic Disorders in the Adult

DISORDER	APPROX. PREVALENCE IN ADULTS^a	SCREENING/TESTING RECOMMENDATIONS^b	CHAPTER(S)
Obesity	36% BMI ≥ 30 70% BMI ≥ 25	Calculate BMI Measure waist circumference Exclude secondary causes Consider comorbid complications	395
Type 2 diabetes mellitus	>8%	Beginning at age 45, screen every 3 years, or earlier in high-risk groups: FPG >126 mg/dL Random plasma glucose >200 mg/dL An elevated HbA _{1c} Consider comorbid complications	396
Hyperlipidemia	20–25%	Cholesterol screening at least every 5 years; more often in high-risk groups Lipoprotein analysis (LDL, HDL) for increased cholesterol, CAD, diabetes Consider secondary causes	400
Metabolic syndrome	35%	Measure waist circumference, FPG, BP, lipids	401
Hypothyroidism	5–10%, women 0.5–2%, men	TSH; confirm with free T ₄ Screen women after age 35 and every 5 years thereafter	377
Graves' disease	1–3%, women 0.1%, men	TSH, free T ₄	376
Thyroid nodules and neoplasia	2–5% palpable >25% by ultrasound	Physical examination of thyroid Fine-needle aspiration biopsy	378

DISORDER	APPROX. PREVALENCE IN ADULTS ^a	SCREENING/TESTING RECOMMENDATIONS ^b	CHAPTER(S)
Osteoporosis	5–10%, women 2–5%, men	Bone mineral density measurements in women >65 years or in postmenopausal women or men at risk Exclude secondary causes	404
Hyperparathyroidism	0.1–0.5%, women > men	Serum calcium PTH, if calcium is elevated Assess comorbid conditions	403
Infertility	10%, couples	Investigate both members of couple Semen analysis in male Assess ovulatory cycles in female Specific tests as indicated	384, 385
Polycystic ovarian syndrome	5–10%, women	Free testosterone, DHEAS Consider comorbid conditions	385
Hirsutism	5–10%	Free testosterone, DHEAS Exclude secondary causes Additional tests as indicated	387
Menopause	Median age, 51	FSH	388
Hyperprolactinemia	15% in women with amenorrhea or galactorrhea	PRL level MRI, if not medication-related	373
Erectile dysfunction	10–25%	Careful history, PRL, testosterone Consider secondary causes (e.g., diabetes)	390
Hypogonadism, male	1–2%	Testosterone, LH	384
Gynecomastia	15%	Often, no tests are indicated Consider Klinefelter's syndrome Consider medications, hypogonadism, liver disease	384

DISORDER	APPROX. PREVALENCE IN ADULTS ^a	SCREENING/TESTING RECOMMENDATIONS ^b	CHAPTER(S)
Klinefelter's syndrome	0.2%, men	Karyotype Testosterone	383
Vitamin D deficiency	10%	Measure serum 25-OH vitamin D Consider secondary causes	402
Turner's syndrome	0.03%, women	Karyotype Consider comorbid conditions	383

^aThe prevalence of most disorders varies among ethnic groups and with aging. Data based primarily on U.S. population. ^bSee individual chapters for additional information on evaluation and treatment. Early testing is indicated in patients with signs and symptoms of disease and in those at increased risk.

Abbreviations: BMI, body mass index; BP, blood pressure; CAD, coronary artery disease; DHEAS, dehydroepiandrosterone; FPG, fasting plasma glucose; FSH, follicle-stimulating hormone; HbA_{1C}, hemoglobin A_{1C}; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LH, luteinizing hormone; MRI, magnetic resonance imaging; PRL, prolactin; PTH, parathyroid hormone; TSH, thyroid-stimulating hormone.

FURTHER READING

Golden SH et al: Prevalence and incidence of endocrine and metabolic disorders in the United States: A comprehensive review. *J Clin Endocrinol Metab* 94:1853, 2009. [[PubMed: 19494161](#)]

Golden SH et al: Health disparities in endocrine disorders: Biological, clinical, and nonclinical factors—an Endocrine Society Scientific Statement. *J Clin Endocrinol Metab* 97:E1579, 2012. [[PubMed: 22730516](#)]

Jameson JL, DeGroot LJ (eds): *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, Elsevier, 2016.

Loriaux DL: *A Biographical History of Endocrinology*. Hoboken, Wiley Blackwell, 2016.

Melmed S, Polonsky KS, Larsen PR, Kronenberg HM (eds): *Williams Textbook of Endocrinology*, 13th ed. Philadelphia, Elsevier, 2016.

3/14/2018

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 370: Mechanisms of Hormone Action

J. Larry Jameson

INTRODUCTION

Hormones function as a communication system within the body. The endocrine system, composed of various glands and the hormones they produce, interacts with essentially all other physiologic systems to regulate growth, metabolism, homeostasis, and reproduction. Because hormones circulate and act via receptors in target tissues, they serve to integrate physiologic responses to external or internal cues. For example, the light-dark cycle, sensed through the visual system, modulates hypothalamic corticotropin-releasing hormone (CRH), which increases pituitary adrenocorticotropin hormone (ACTH) production, leading to increased adrenal cortisol production before the time of waking in the morning. Increased cortisol, in turn, circulates throughout the body, acting via the nuclear glucocorticoid receptor, to activate numerous genetic programs that influence metabolism, the cardiovascular system, behavior, and the immune system. This chapter provides an overview of the different types of hormones and how they function at the cellular level to control myriad physiologic processes.

CLASSES OF HORMONES

Hormones can be divided into five major types: (1) *amino acid derivatives* such as dopamine, catecholamine, and thyroid hormone; (2) *small neuropeptides* such as gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), somatostatin, and vasopressin; (3) *large proteins* such as insulin, luteinizing hormone (LH), and parathyroid hormone (PTH); (4) *steroid hormones* such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) *vitamin derivatives* such as retinoids ([vitamin A](#)) and [vitamin D](#). A variety of *peptide growth factors*, most of which act locally, share actions with hormones. As a rule, amino acid derivatives and peptide hormones interact with cell-surface membrane receptors. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and interact with intracellular nuclear receptors, although many also interact with membrane receptors or intracellular signaling proteins as well.

HORMONE AND RECEPTOR FAMILIES

Hormones and receptors can be grouped into families, reflecting structural similarities and evolutionary origins ([Table 370-1](#)). The evolution of these families generates diverse but highly selective pathways of

hormone action. Recognition of these relationships has proven useful for extrapolating information gleaned from one hormone or receptor to other family members.

TABLE 370-1

Examples of Membrane Receptor Families and Signaling Pathways

RECEPTORS	EFFECTORS	SIGNALING PATHWAYS
G Protein–Coupled Seven-Transmembrane Receptor (GPCR)		
β -Adrenergic, LH, FSH, TSH Glucagon, PTH, PTHrP, ACTH, MSH, GHRH, CRH α -Adrenergic, somatostatin TRH, GnRH	$G_s\alpha$, adenylate cyclase Ca^{2+} channels $G_i\alpha$ G_q, G_{11}	Stimulation of cyclic AMP production, protein kinase A Calmodulin, Ca^{2+} -dependent kinases Inhibition of cyclic AMP production Activation of K^+ , Ca^{2+} channels Phospholipase C, diacyl-glycerol, IP_3 , protein kinase C, voltage-dependent Ca^{2+} channels
Receptor Tyrosine Kinase		
Insulin, IGF-I EGF, NGF	Tyrosine kinases, IRS Tyrosine kinases, ras	MAP kinases, PI 3-kinase; AKT Raf, MAP kinases, RSK
Cytokine Receptor–Linked Kinase		
GH, PRL	JAK, tyrosine kinases	STAT, MAP kinase, PI 3-kinase, IRS-1
Serine Kinase		
Activin, TGF- β , MIS	Serine kinase	Smads

Abbreviations: IP_3 , inositol triphosphate; IRS, insulin receptor substrates; MAP, mitogen-activated protein; MSH, melanocyte-stimulating hormone; NGF, nerve growth factor; PI, phosphatidylinositol; RSK, ribosomal S6 kinase; TGF- β , transforming growth factor β . For all other abbreviations, see text. Note that most receptors interact with multiple effectors and activate networks of signaling pathways.

The glycoprotein hormone family, consisting of thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), LH, and human chorionic gonadotropin (hCG), illustrates many features of evolutionarily related hormones. The glycoprotein hormones are heterodimers that share the α subunit in common; the β subunits are distinct and confer specific biologic actions. The overall three-dimensional architecture of the β subunits is similar, reflecting the locations of conserved disulfide bonds that restrain protein conformation. The cloning of the β -subunit genes from multiple species suggests that this family arose from a common ancestral gene, probably by gene duplication and subsequent divergence to evolve new biologic functions.

As hormone families enlarge and diverge, their receptors must co-evolve to derive new biologic functions. Related G protein-coupled receptors (GPCRs), for example, have evolved for each of the glycoprotein hormones. These receptors are also structurally similar, and each is coupled predominantly to the $G_{s\alpha}$ signaling pathway. However, there is minimal overlap of hormone binding. For example, TSH binds with high specificity to the TSH receptor but interacts minimally with the LH or FSH receptors. Nonetheless, there can be subtle physiologic consequences of hormone cross-reactivity with other receptors. Very high levels of hCG during pregnancy stimulate the TSH receptor and increase thyroid hormone levels, resulting via feedback inhibition in a compensatory decrease in TSH.

Insulin and insulin-like growth factor I (IGF-I) and IGF-II have structural similarities that are most apparent when precursor forms of the proteins are compared. In contrast to the high degree of specificity seen with the glycoprotein hormones, there is moderate cross-talk among the members of the insulin/IGF family. High concentrations of an IGF-II precursor produced by certain tumors (e.g., sarcomas) can cause hypoglycemia, partly because of binding to insulin and IGF-I receptors (**Chap. 403**). High concentrations of insulin also bind to the IGF-I receptor, perhaps accounting for some of the clinical manifestations seen in conditions with chronic hyperinsulinemia.

Another important example of receptor cross-talk is seen with PTH and parathyroid hormone-related peptide (PTHrP) (**Chap. 403**). PTH is produced by the parathyroid glands, whereas PTHrP is expressed at high levels during development and by a variety of tumors (**Chap. 89**). These hormones have amino acid sequence similarity, particularly in their amino-terminal regions. Both hormones bind to the PTH1R receptor that is expressed in bone and kidney. Hypercalcemia and hypophosphatemia therefore may result from excessive production of either hormone, making it difficult to distinguish hyperparathyroidism from hypercalcemia of malignancy solely on the basis of serum chemistries. However, sensitive and specific assays for PTH and PTHrP now allow these disorders to be distinguished more readily.

Based on their specificities for DNA-binding sites, the nuclear receptor family can be subdivided into type 1 receptors (glucocorticoid receptor, mineralocorticoid receptor, androgen receptor, estrogen receptor, progesterone receptor) that bind steroids and type 2 receptors (thyroid hormone receptor, vitamin D receptor, retinoic acid receptor, peroxisome proliferator activated receptor) that bind thyroid hormone, vitamin D, retinoic acid, or lipid derivatives, respectively. Certain functional domains in nuclear receptors, such as the zinc finger DNA-binding domains, are highly conserved. However, selective amino acid differences within this domain confer DNA sequence specificity. The hormone-binding domains are more

variable, providing great diversity in the array of small molecules that bind to different nuclear receptors. With few exceptions, hormone binding is highly specific for a single type of nuclear receptor. One exception involves the glucocorticoid and mineralocorticoid receptors. Because the mineralocorticoid receptor also binds glucocorticoids with high affinity, an enzyme (11 β -hydroxysteroid dehydrogenase) in renal tubular cells inactivates glucocorticoids, allowing selective responses to mineralocorticoids such as aldosterone. However, when very high glucocorticoid concentrations occur, as in Cushing's syndrome, the glucocorticoid degradation pathway becomes saturated, allowing excessive cortisol levels to bind mineralocorticoid receptors leading to sodium retention and potassium wasting. This phenomenon is particularly pronounced in ectopic adrenocorticotrophic hormone (ACTH) syndromes ([Chap. 379](#)). Another example of relaxed nuclear receptor specificity involves the estrogen receptor, which can bind an array of compounds, some of which have little apparent structural similarity to the high-affinity ligand [estradiol](#). This feature of the estrogen receptor makes it susceptible to activation by “environmental [estrogens](#)” such as resveratrol, octylphenol, and many other aromatic hydrocarbons. However, this lack of specificity provides an opportunity to synthesize a remarkable series of clinically useful antagonists (e.g., tamoxifen) and selective estrogen response modulators (SERMs) such as raloxifene. These compounds generate distinct conformations that alter receptor interactions with components of the transcription machinery (see below), thereby conferring their unique actions.

HORMONE SYNTHESIS AND PROCESSING

The synthesis of peptide hormones and their receptors occurs through a classic pathway of gene expression: transcription \rightarrow mRNA \rightarrow protein \rightarrow posttranslational protein processing \rightarrow intracellular sorting, followed by membrane integration or secretion.

Many hormones are embedded within larger precursor polypeptides that are proteolytically processed to yield the biologically active hormone. Examples include proopiomelanocortin (POMC) \rightarrow ACTH; proglucagon \rightarrow glucagon; proinsulin \rightarrow insulin; and pro-PTH \rightarrow PTH, among others. In many cases, such as POMC and proglucagon, these precursors generate multiple biologically active peptides. It is provocative that hormone precursors are typically inactive, presumably adding an additional level of regulatory control. Prohormone conversion occurs not only for peptide hormones but also for certain steroids (testosterone \rightarrow dihydrotestosterone) and thyroid hormone ($T_4 \rightarrow T_3$).

Peptide precursor processing is intimately linked to intracellular sorting pathways that transport proteins to appropriate vesicles and enzymes, resulting in specific cleavage steps, followed by protein folding and translocation to secretory vesicles. Hormones destined for secretion are translocated across the endoplasmic reticulum under the guidance of an amino-terminal signal sequence that subsequently is cleaved. Cell-surface receptors are inserted into the membrane via short segments of hydrophobic amino acids that remain embedded within the lipid bilayer. During translocation through the Golgi and endoplasmic reticulum, hormones and receptors are subject to a variety of posttranslational modifications, such as glycosylation and phosphorylation, which can alter protein conformation, modify circulating half-life, and alter biologic activity.

Synthesis of most steroid hormones is based on modifications of the precursor, cholesterol. Multiple regulated enzymatic steps are required for the synthesis of testosterone (**Chap. 384**), estradiol (**Chap. 385**), cortisol (**Chap. 379**), and vitamin D (**Chap. 402**). This large number of synthetic steps predisposes to multiple genetic and acquired disorders of steroidogenesis.

Endocrine genes contain regulatory DNA elements similar to those found in many other genes, but their exquisite control by hormones reflects the presence of specific hormone response elements. For example, the TSH genes are repressed directly by thyroid hormones acting through the thyroid hormone receptor (TR), a member of the nuclear receptor family. Steroidogenic enzyme gene expression requires specific transcription factors, such as steroidogenic factor-1 (SF-1), acting in conjunction with signals transmitted by trophic hormones (e.g., ACTH or LH). Once activated, SF-1 functions as a master regulator, inducing a large array of genes required for steroidogenic and metabolic pathways required for steroid synthesis. For some hormones, substantial regulation occurs at the level of translational efficiency. Insulin biosynthesis, although it requires ongoing gene transcription, is regulated primarily at the translational and secretory levels in response to elevated levels of glucose or amino acids.

HORMONE SECRETION, TRANSPORT, AND DEGRADATION

The circulating level of a hormone is determined by its rate of secretion and its half-life. After protein processing, peptide hormones (e.g., GnRH, insulin, growth hormone [GH]) are stored in secretory granules. As these granules mature, they are poised beneath the plasma membrane for imminent release into the circulation. In most instances, the stimulus for hormone secretion is a releasing factor or neural signal that induces rapid changes in intracellular calcium concentrations, leading to secretory granule fusion with the plasma membrane and release of its contents into the extracellular environment and bloodstream. Steroid hormones, in contrast, diffuse into the circulation as they are synthesized. Thus, their secretory rates are closely aligned with rates of synthesis. For example, ACTH and LH induce steroidogenesis by stimulating the activity of the *steroidogenic acute regulatory* (StAR) protein (transports cholesterol into the mitochondrion) along with other rate-limiting steps (e.g., cholesterol side-chain cleavage enzyme, CYP11A1) in the steroidogenic pathway.

Hormone transport and degradation dictate the rapidity with which a hormonal signal decays. Some hormone signals are evanescent (e.g., somatostatin), whereas others are longer-lived (e.g., TSH). Because somatostatin exerts effects in virtually every tissue, a short half-life allows its concentrations and actions to be controlled locally. Structural modifications that impair somatostatin degradation have been useful for generating long-acting therapeutic analogues such as octreotide (**Chap. 373**). In contrast, the actions of TSH are highly specific for the thyroid gland. Its prolonged half-life accounts for relatively constant serum levels even though TSH is secreted in discrete pulses.

An understanding of circulating hormone half-life is important for achieving physiologic hormone replacement, as the frequency of dosing and the time required to reach steady state are intimately linked to rates of hormone decay. T₄, for example, has a circulating half-life of 7 days. Consequently, >1 month is

required to reach a new steady state, and single daily doses are sufficient to achieve constant hormone levels. T_3 , in contrast, has a half-life of 1 day. Its administration is associated with more dynamic serum levels, and it must be administered two to three times per day. Similarly, synthetic glucocorticoids vary widely in their half-lives; those with longer half-lives (e.g., [dexamethasone](#)) are associated with greater suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Most protein hormones (e.g., ACTH, GH, prolactin [PRL], PTH, LH) have relatively short half-lives (<20 min), leading to sharp peaks of secretion and decay. The only accurate way to profile the pulse frequency and amplitude of these hormones is to measure levels in frequently sampled blood (every 10 min or less) over long durations (8–24 h). Because this is not practical in a clinical setting, an alternative strategy is to pool three to four samples drawn at about 30-min intervals, or interpret the results in the context of a relatively wide normal range. Rapid hormone decay is useful in certain clinical settings. For example, the short half-life of PTH allows the use of intraoperative PTH determinations to confirm successful removal of a parathyroid adenoma. This is particularly valuable diagnostically when there is a possibility of multicentric disease or parathyroid hyperplasia, as occurs with multiple endocrine neoplasia (MEN) or renal insufficiency.

Many hormones circulate in association with serum-binding proteins. Examples include (1) T_4 and T_3 binding to thyroxine-binding globulin (TBG), albumin, and thyroxine-binding prealbumin (TBPA); (2) cortisol binding to cortisol-binding globulin (CBG); (3) androgen and estrogen binding to sex hormone-binding globulin (SHBG); (4) IGF-I and II binding to multiple IGF-binding proteins (IGFBPs); (5) GH interactions with GH-binding protein (GHBP), a circulating fragment of the GH receptor extracellular domain; and (6) activin binding to follistatin. These interactions provide a hormonal reservoir, prevent otherwise rapid degradation of unbound hormones, restrict hormone access to certain sites (e.g., IGFBPs), and modulate the unbound, or “free,” hormone concentrations. Although a variety of binding protein abnormalities have been identified, most have little clinical consequence aside from creating diagnostic problems. For example, TBG deficiency can reduce total thyroid hormone levels greatly but the free concentrations of T_4 and T_3 remain normal. Liver disease and certain medications can also influence binding protein levels (e.g., estrogen increases TBG) or cause displacement of hormones from binding proteins (e.g., salicylate displaces T_4 from TBG). In general, only unbound hormone is available to interact with receptors and thus elicit a biologic response. Short-term perturbations in binding proteins change the free hormone concentration, which in turn induces compensatory adaptations through feedback loops. SHBG changes in women are an exception to this self-correcting mechanism. When SHBG decreases because of insulin resistance or androgen excess, the unbound testosterone concentration is increased, potentially contributing to hirsutism in women with polycystic ovary syndrome (PCOS) ([Chap. 387](#)). The increased unbound testosterone level does not result in an adequate compensatory feedback correction because estrogen, not testosterone, is the primary regulator of the reproductive axis.

An additional exception to the unbound hormone hypothesis involves megalin, a member of the low-density lipoprotein (LDL) receptor family that serves as an endocytotic receptor for thyroglobulin, carrier-bound vitamins A and D and SHBG-bound androgens and [estrogens](#). After internalization, the carrier proteins are

degraded in lysosomes and release their bound ligands within the cells. Membrane transporters have also been identified for thyroid hormones.

Hormone degradation can be an important mechanism for regulating concentrations locally. As noted above, 11 β -hydroxysteroid dehydrogenase inactivates glucocorticoids in renal tubular cells, preventing actions through the mineralocorticoid receptor. Thyroid hormone deiodinases convert T₄ to T₃ and can inactivate T₃. During development, degradation of retinoic acid by Cyp26b1 prevents primordial germ cells in the male from entering meiosis, as occurs in the female ovary.

HORMONE ACTION THROUGH RECEPTORS

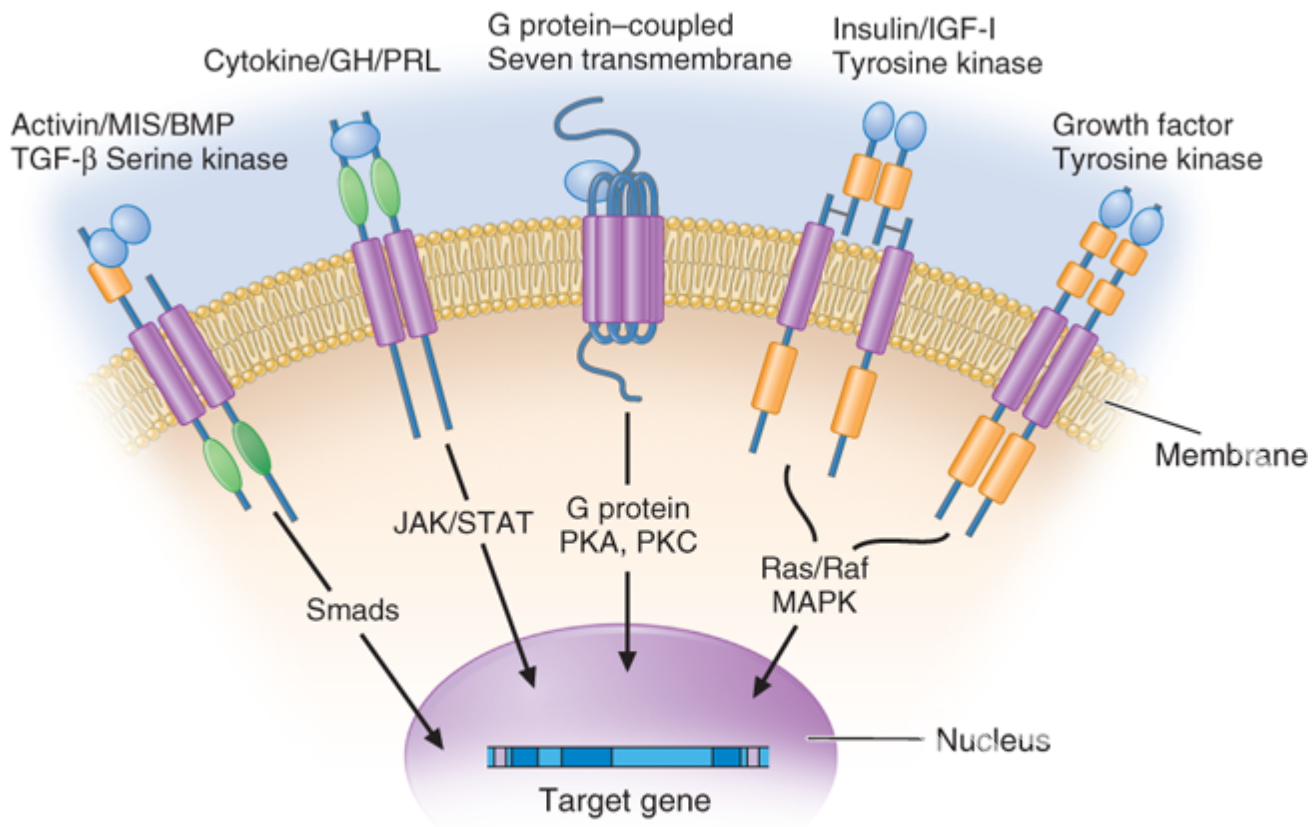
Receptors for hormones are divided into two major classes: membrane and nuclear. *Membrane receptors* primarily bind peptide hormones and catecholamines. *Nuclear receptors* bind small molecules that can diffuse across the cell membrane, such as steroids and vitamin D. Certain general principles apply to hormone-receptor interactions regardless of the class of receptor. Hormones bind to receptors with specificity and an affinity that generally coincides with the dynamic range of circulating hormone concentrations. Low concentrations of free hormone (usually 10⁻¹² to 10⁻⁹ M) rapidly associate and dissociate from receptors in a bimolecular reaction such that the occupancy of the receptor at any given moment is a function of hormone concentration and the receptor's affinity for the hormone. Receptor numbers vary greatly in different target tissues, providing one of the major determinants of specific tissue responses to circulating hormones. For example, ACTH receptors are located almost exclusively in the adrenal cortex, and FSH receptors are found predominantly in the gonads. In contrast, insulin and TRs are widely distributed, reflecting the need for metabolic responses in all tissues.

MEMBRANE RECEPTORS

Membrane receptors for hormones can be divided into several major groups: (1) seven transmembrane GPCRs, (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) serine kinase receptors (**Fig. 370-1**). The *seven transmembrane GPCR family* binds a remarkable array of hormones, including large proteins (e.g., LH, PTH), small peptides (e.g., TRH, somatostatin), catecholamines (**epinephrine**, dopamine), and even minerals (e.g., calcium). The extracellular domains of GPCRs vary widely in size and are the major binding site for large hormones. The transmembrane-spanning regions are composed of hydrophobic α -helical domains that traverse the lipid bilayer. Like some channels, these domains are thought to circularize and form a hydrophobic pocket into which certain small ligands fit. Hormone binding induces conformational changes in these domains, transducing structural changes to the intracellular domain, which is a docking site for G proteins.

FIGURE 370-1

Membrane receptor signaling. MAPK, mitogen-activated protein kinase; PKA, C, protein kinase A, C; TGF, transforming growth factor. For other abbreviations, see text.



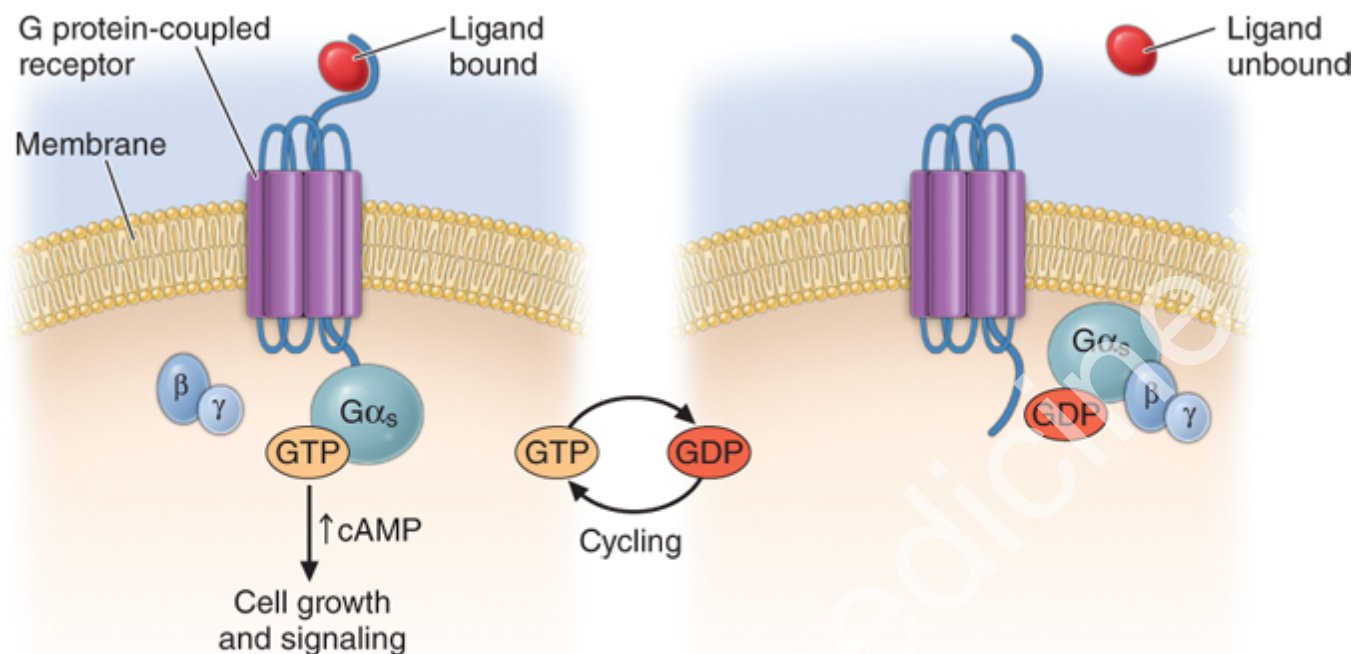
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The large family of *G proteins*, so named because they bind guanine nucleotides (guanosine triphosphate [GTP], guanosine diphosphate [GDP]), provides great diversity for coupling receptors to different signaling pathways. G proteins form a heterotrimeric complex that is composed of various α and $\beta\gamma$ subunits (**Fig. 370-2**). The α subunit contains the guanine nucleotide-binding site and an intrinsic GTPase that hydrolyzes GTP \rightarrow GDP. The $\beta\gamma$ subunits are tightly associated and modulate the activity of the α subunit as well as mediating their own effector signaling pathways. G protein activity is regulated by a cycle that involves GTP hydrolysis and dynamic interactions between the α and $\beta\gamma$ subunits. Hormone binding to the receptor induces GDP dissociation, allowing $G\alpha$ to bind GTP and dissociate from the $\beta\gamma$ complex. Under these conditions, the $G\alpha$ subunit is activated and mediates signal transduction through various enzymes, such as adenylate cyclase and phospholipase C. GTP hydrolysis to GDP allows reassociation with the $\beta\gamma$ subunits and restores the inactive state. G proteins interact with other cellular proteins, including kinases, channels, G protein-coupled receptor kinases (GRKs), and arrestins, that mediate signaling as well as receptor desensitization and recycling.

FIGURE 370-2

G protein signaling. G protein-coupled receptors signal via the family of G proteins, so-named because they bind guanylyl nucleotides. In the example shown, a G-protein-coupled receptor (GPCR) bound to a ligand induces GDP dissociation, allowing $G\alpha$ to bind GTP and dissociate from the $\beta\gamma$ complex. GTP-bound $G\alpha$ increases cAMP production by adenyl cyclase and activates the protein kinase A pathway. Not shown are separate signaling pathways activated by the $\beta\gamma$ complex. When GTP is converted to GDP by an intrinsic GTPase, the $\beta\gamma$ subunits reassociate with GDP-bound $G\alpha$ and the complex returns to an inactive state. As

noted in the text, mutations in Gs α that eliminate GTPase activity result in constitutive activation of receptor signaling pathways because GTP-bound Gs α cannot be converted to its GDP-bound inactive state. cAMP, cyclic adenosine 5'-monophosphate; GDP, guanosine diphosphate; Gs α , G protein α ; GTP, guanosine triphosphate.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

A variety of endocrinopathies result from mutations in GPCR receptors that alter their interactions with G proteins (**Table 370-2**). Loss-of-function mutations are generally recessive and inactivate the relevant hormone signaling pathway. Because many of these receptors are important for development as well as signaling, patient presentations often resemble glandular failure syndromes (e.g., mutations in LH-R, FSH-R, TSH-R). Gain-of-function mutations involve a more complex mechanism. Selected mutations induce conformational changes in the GPCR that mimic the activated state normally induced by hormone binding. These mutations result in a constitutively active state in which G protein coupling stimulates cell signaling pathways, most commonly via cyclic adenosine 5'-monophosphate (cAMP) and protein kinase A. When mutations occur in the germline, the conditions are heritable and present in early life (e.g., LH-R, TSH-R). Somatic mutations can also occur and result in clonal expansion of hyperfunctioning cells.

TABLE 370-2

Genetic Causes of G protein Receptor Disorders

RECEPTOR	DISORDER	GENETICS
LH	Leydig cell hypoplasia (male) Primary amenorrhea, resistance to LH (female) Familial male precocious puberty (male) Leydig cell adenoma, precocious puberty (male)	AR, inactivating AR, inactivating AD, activating Sporadic, activating
FSH	Hypergonadotropic ovarian failure (female) Hypospermia (male) Ovarian hyperstimulation (female)	AR, inactivating AR, inactivating AD, activating
TSH	Congenital hypothyroidism, TSH resistance Nonautoimmune familial hyperthyroidism Hyperfunctioning thyroid adenoma	AR, AD, inactivating AD, activating Sporadic, activating
GnRH	Hypogonadotropic hypogonadism	AR, inactivating
Kisspeptin	Hypogonadotropic hypogonadism Precocious puberty	AR, inactivating AD, activating
TRH	Central hypothyroidism	AR, inactivating
GHRH	GH deficiency	AR, inactivating
PTH	Blomstrand chondrodysplasia Jansen metaphyseal chondrodysplasia	AR, inactivating AD, activating
Calcium sensing receptor	Familial hypocalciuric hypercalcemia Neonatal severe hyperparathyroidism Familial hypocalcemic hypercalciura	AD, inactivating AR, inactivating AD, activating
Arginine vasopressin receptor 2	Nephrogenic diabetes insipidus Nephrogenic SIADH	XL, inactivating XL, activating
ACTH	Familial ACTH resistance	AR, inactivating

RECEPTOR	DISORDER	GENETICS
Melanocortin 4	Severe obesity	Codominant, inactivating

Abbreviations: ACTH, adrenocorticotropin hormone; AD, autosomal dominant; AR, autosomal recessive; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone–releasing hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone secretion; TRH, thyrotropin-releasing hormone; TSH, thyroid-stimulating hormone; XL, X-linked.

Mutations in the TSH-R illustrate the range of possible clinical consequences of GPCR mutations. Recessive inactivating mutations in the TSH-R cause congenital hypothyroidism with thyroid gland hypoplasia and resistance to TSH. Clinically, the hormone profile resembles primary hypothyroidism with low T4 and high TSH. On the other hand, germline activating mutations cause congenital hyperthyroidism. The disorder is autosomal dominant because an activating mutation of one TSH-R allele is sufficient to induce cellular hyperfunction and disease. Because the TSH-R is activated in every cell of the thyroid, there is hyperplastic growth and hyperfunction that resembles the pathology seen in Graves' disease. This unusual disorder presents in infancy and must be distinguished from the more common clinical circumstance in which maternal antibodies in women with active or previously treated Graves' disease cross the placenta and stimulate the thyroid gland of the fetus. If an activating TSH-R mutation occurs later in life, in the somatic tissue, there is clonal expansion of the thyrocyte harboring the mutation, ultimately leading to an autonomous hyperfunctioning thyroid nodule. Of note, a similar condition can be caused by somatic mutations in G_{α} . In this case, the G_{α} GTPase is inactivated and GTP cannot be converted to GDP. Consequently, the G_{α} signaling pathway in this particular cell is constitutively active, mimicking chronic TSH stimulation and again leading to clonal expansion and an autonomous hyperfunctioning thyroid nodule. About one-third of hyperfunctioning "hot" thyroid nodules harbor mutations in either the TSH-R or G_{α} (TSH-R mutations are more common).

G_{α} mutations in tissues other than the thyroid can also cause endocrine disease. For example, G_{α} mutations in pituitary somatotropes mimic activation of the growth hormone–releasing hormone (GHRH) pathway and lead to GH-producing adenomas and acromegaly. Rarely, mutations in other components of the protein kinase A pathway in somatotropes can also cause GH-producing adenomas. G_{α} mutations that occur early in development (typically mosaic) cause McCune-Albright syndrome ([Chap. 405](#)) and the clinical features are manifest because the activated G protein pathway mimics the actions of various hormones [PTH, melanocyte stimulating hormone (MSH), TSH, GHRH] in different tissues. Germline inactivating G_{α} mutations cause a range of disorders that are transmitted and expressed in a complex manner because the locus is imprinted ([Chap. 403](#)). These conditions include Albright's hereditary osteodystrophy (AHO), pseudopseudohypoparathyroidism (PPHP), and pseudohypoparathyroidism types 1b, 1c, and 2.

The *tyrosine kinase receptors* transduce signals for insulin and a variety of growth factors, such as IGF-I, epidermal growth factor (EGF), nerve growth factor, platelet-derived growth factor, and fibroblast growth factor. The cysteine-rich extracellular domains contain binding sites for the growth factors. After ligand binding, this class of receptors undergoes autophosphorylation, inducing interactions with intracellular adaptor proteins such as Shc and insulin receptor substrates (IRS). In the case of the insulin receptor, multiple kinases are activated, including the Raf-Ras-MAPK and the Akt/protein kinase B pathways. The tyrosine kinase receptors play a prominent role in cell growth and differentiation as well as in intermediary metabolism.

The GH and PRL receptors belong to the *cytokine receptor* family. Analogous to the tyrosine kinase receptors, ligand binding induces receptor interaction with intracellular kinases—the Janus kinases (JAKs), which phosphorylate members of the signal transduction and activators of transcription (STAT) family—as well as with other signaling pathways (Ras, PI3-K, MAPK). The activated STAT proteins translocate to the nucleus and stimulate expression of target genes.

The *serine kinase receptors* mediate the actions of activins, transforming growth factor β , müllerian-inhibiting substance (MIS, also known as anti-müllerian hormone, AMH), and bone morphogenic proteins (BMPs). This family of receptors (consisting of type I and II subunits) signals through proteins termed *smads* (fusion of terms for *Caenorhabditis elegans* sma + mammalian mad). Like the STAT proteins, the smads serve a dual role of transducing the receptor signal and acting as transcription factors. The pleomorphic actions of these growth factors dictate that they act primarily in a local (paracrine or autocrine) manner. Binding proteins such as follistatin (which binds activin and other members of this family) function to inactivate the growth factors and restrict their distribution.

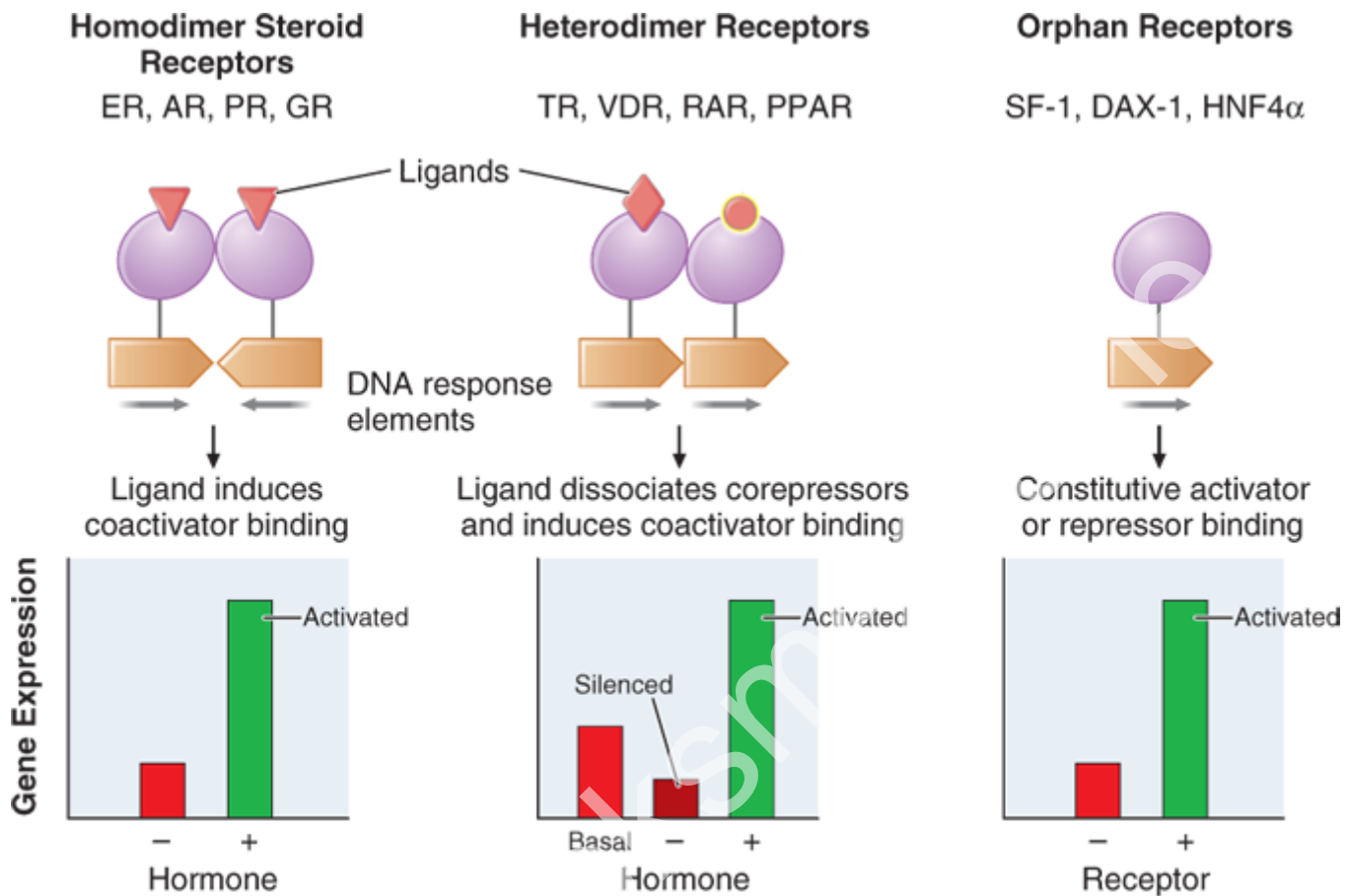
Disease-causing mutations also occur in each of these classes of receptors. For example, insulin receptor mutations cause an extreme form of insulin resistance. GH receptor mutations cause Laron-type dwarfism, characterized by low IGF-1 and high GH. AMH receptor mutations cause persistent Müllerian duct syndrome. These hormone resistance syndromes are autosomal recessive and relatively uncommon. Unlike the GPCRs, activating mutations are unusual, although they do occur for the RET tyrosine kinase receptor, which causes the autosomal dominant disorder multiple endocrine neoplasia type 2 (MEN-2) ([Chap. 381](#)).

NUCLEAR RECEPTORS

The family of nuclear receptors has grown to nearly 100 members, many of which are still classified as orphan receptors because their ligands, if they exist, have not been identified ([Fig. 370-3](#)). Otherwise, most nuclear receptors are classified on the basis of their ligands. Although all nuclear receptors ultimately act to increase or decrease gene transcription, some (e.g., glucocorticoid receptor) reside primarily in the cytoplasm, whereas others (e.g., TR) are located in the nucleus. After ligand binding, the cytoplasmically localized receptors translocate to the nucleus. There is growing evidence that certain nuclear receptors (e.g., glucocorticoid, estrogen) can also act at the membrane or in the cytoplasm to activate or repress signal transduction pathways, providing a mechanism for cross-talk between membrane and nuclear receptors.

FIGURE 370-3

Nuclear receptor signaling. AR, androgen receptor; DAX, *dosage-sensitive sex-reversal, adrenal hypoplasia congenita, X-chromosome*; ER, estrogen receptor; GR, glucocorticoid receptor; HNF4 α , hepatic nuclear factor 4 α ; PPAR, peroxisome proliferator activated receptor; PR, progesterone receptor; RAR, retinoic acid receptor; SF-1, steroidogenic factor-1; TR, thyroid hormone receptor; VDR, vitamin D receptor.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

The structures of nuclear receptors have been studied extensively, including by x-ray crystallography. The DNA-binding domain, consisting of two zinc fingers, contacts-specific DNA recognition sequences in target genes. Most nuclear receptors bind to DNA as dimers. Consequently, each monomer recognizes an individual DNA motif, referred to as a "half-site." The steroid receptors, including the glucocorticoid, estrogen, progesterone, and androgen receptors, bind to DNA as homodimers. Consistent with this twofold symmetry, their DNA recognition half-sites are palindromic. The thyroid, retinoid, peroxisome proliferator activated, and vitamin D receptors bind to DNA preferentially as heterodimers in combination with retinoid X receptors (RXRs). Their DNA half-sites are typically arranged as direct repeats.

The carboxy-terminal hormone-binding domain mediates transcriptional control. For type II receptors such as TR and retinoic acid receptor (RAR), co-repressor proteins bind to the receptor in the absence of ligand and silence gene transcription. Hormone binding induces conformational changes, triggering the release of co-repressors and inducing the recruitment of coactivators that stimulate transcription. Thus, these receptors are capable of mediating dramatic changes in the level of gene activity. Disease states can be associated with defective regulation of these events. For example, in promyelocytic leukemia, fusion of RAR α to other nuclear

proteins causes aberrant gene silencing that prevents normal cellular differentiation. Treatment with retinoic acid reverses this repression and allows cellular differentiation and apoptosis to occur. Most type 1 steroid receptors interact weakly with co-repressors, but ligand binding still induces interactions with an array of coactivators. X-ray crystallography shows that various SERMs induce distinct estrogen receptor conformations. The tissue-specific responses caused by these agents in breast, bone, and uterus appear to reflect distinct interactions with coactivators. The receptor-coactivator complex stimulates gene transcription by several pathways, including (1) recruitment of enzymes (histone acetyl transferases) that modify chromatin structure, (2) interactions with additional transcription factors on the target gene, and (3) direct interactions with components of the general transcription apparatus to enhance the rate of RNA polymerase II-mediated transcription. Studies of nuclear receptor-mediated transcription show that these are dynamic events that involve relatively rapid (e.g., 30–60 min) cycling of transcription complexes on any specific target gene.

Nuclear receptor mutations are an important cause of endocrine disease. Androgen receptor mutations cause androgen insensitivity syndrome (AIS) ([Chap. 383](#)). Because the androgen receptor is located on the X-chromosome, mutations are more commonly manifest than with other nuclear receptor disorders. Affected individuals with AIS are XY phenotypic females with retained testes and male-range testosterone levels. Tissue insensitivity to androgens varies based on the severity of the mutation. Müllerian structures are absent because Sertoli cells of the testis produce AMH during development. Female carriers of androgen receptor mutations are phenotypically normal. Recessive mutations of the estrogen, glucocorticoid, and vitamin D receptors are rare.

Thyroid hormone receptor β (TR β) mutations have an unusual pathophysiology. They are autosomal dominant and function via a “dominant negative” mechanism to cause resistance to thyroid hormone (RTH) ([Chap. 375](#)). The mutations occur in selected regions of the TR β hormone-binding domain and preserve the ability of the mutant receptor to heterodimerize with RXR and bind to DNA regulatory sites. The mutant receptors function as antagonists of receptors from the normal copy of the TR β gene. Affected patients have high T₄ and T₃ and inappropriately elevated (unsuppressed) TSH, reflecting impaired feedback regulation of the hypothalamic-pituitary-thyroid axis. Organ systems are variably resistant to thyroid hormones based upon the relative expression of TR β and TR α proteins. Mutations in the genes encoding TR α and PPAR γ can also cause disease by functioning in an analogous dominant negative manner.

FUNCTIONS OF HORMONES

The functions of individual hormones are described in detail in subsequent chapters. Nevertheless, it is useful to illustrate how most biologic responses require integration of several different hormone pathways. The physiologic functions of hormones can be divided into three general areas: (1) growth and differentiation, (2) maintenance of homeostasis, and (3) reproduction.

GROWTH

Multiple hormones and nutritional factors mediate the complex phenomenon of growth (**Chap. 371**). Short stature may be caused by GH deficiency, hypothyroidism, Cushing's syndrome, precocious puberty, malnutrition, chronic illness, or genetic abnormalities that affect the epiphyseal growth plates (e.g., *FGFR3* and *SHOX* mutations). Many factors (GH, IGF-I, thyroid hormones) stimulate growth, whereas others (sex steroids) lead to epiphyseal closure. Understanding these hormonal interactions is important in the diagnosis and management of growth disorders. For example, delaying exposure to high levels of sex steroids may enhance the efficacy of GH treatment.

MAINTENANCE OF HOMEOSTASIS

Although virtually all hormones affect homeostasis, the most important among them are the following:

1. Thyroid hormone—controls about 25% of basal metabolism in most tissues.
2. Cortisol—exerts a permissive action for many hormones in addition to its own direct effects.
3. PTH—regulates calcium and phosphorus levels.
4. Vasopressin—regulates serum osmolality by controlling renal free-water clearance.
5. Mineralocorticoids—control vascular volume and serum electrolyte (Na^+ , K^+) concentrations.
6. Insulin—maintains euglycemia in the fed and fasted states.

The defense against hypoglycemia is an impressive example of integrated hormone action (**Chap. 399**). In response to the fasting state and falling blood glucose, insulin secretion is suppressed, resulting in decreased glucose uptake and enhanced glycogenolysis, lipolysis, proteolysis, and gluconeogenesis to mobilize fuel sources. If hypoglycemia develops (usually from insulin administration or sulfonylureas), an orchestrated counterregulatory response occurs—glucagon and **epinephrine** rapidly stimulate glycogenolysis and gluconeogenesis, whereas GH and cortisol act over several hours to raise glucose levels and antagonize insulin action.

Although free-water clearance is controlled primarily by vasopressin, cortisol and thyroid hormone are also important for facilitating renal tubular responses to vasopressin (**Chap. 374**). PTH and vitamin D function in an interdependent manner to control calcium metabolism (**Chap. 402**). PTH stimulates renal synthesis of 1,25-dihydroxyvitamin D, which increases calcium absorption in the gastrointestinal tract and enhances PTH action in bone. Increased calcium, along with vitamin D, feeds back to suppress PTH, thus maintaining calcium balance.

Depending on the severity of a specific stress and whether it is acute or chronic, multiple endocrine and cytokine pathways are activated to mount an appropriate physiologic response. In severe acute stress such as trauma or shock, the sympathetic nervous system is activated and catecholamines are released, leading to increased cardiac output and a primed musculoskeletal system. Catecholamines also increase mean blood pressure and stimulate glucose production. Multiple stress-induced pathways converge on the

hypothalamus, stimulating several hormones, including vasopressin and corticotropin-releasing hormone (CRH). These hormones, in addition to cytokines (tumor necrosis factor α , interleukin [IL] 2, IL-6) increase ACTH and GH production. ACTH stimulates the adrenal gland, increasing cortisol, which in turn helps sustain blood pressure and dampen the inflammatory response. Increased vasopressin acts to conserve free water.

REPRODUCTION

The stages of reproduction include (1) sex determination during fetal development (**Chap. 383**); (2) sexual maturation during puberty (**Chaps. 384 and 385**); (3) conception, pregnancy, lactation, and child rearing (**Chap. 385**); and (4) cessation of reproductive capability at menopause (**Chap. 388**). Each of these stages involves an orchestrated interplay of multiple hormones, a phenomenon well illustrated by the dynamic hormonal changes that occur during each 28-day menstrual cycle. In the early follicular phase, pulsatile secretion of LH and FSH stimulates the progressive maturation of the ovarian follicle. This results in gradually increasing estrogen and progesterone levels, leading to enhanced pituitary sensitivity to GnRH, which, when combined with accelerated GnRH secretion, triggers the LH surge and rupture of the mature follicle. Inhibin, a protein produced by the granulosa cells, enhances follicular growth and feeds back to the pituitary to selectively suppress FSH without affecting LH. Growth factors such as EGF and IGF-I modulate follicular responsiveness to gonadotropins. Vascular endothelial growth factor and prostaglandins play a role in follicle vascularization and rupture.

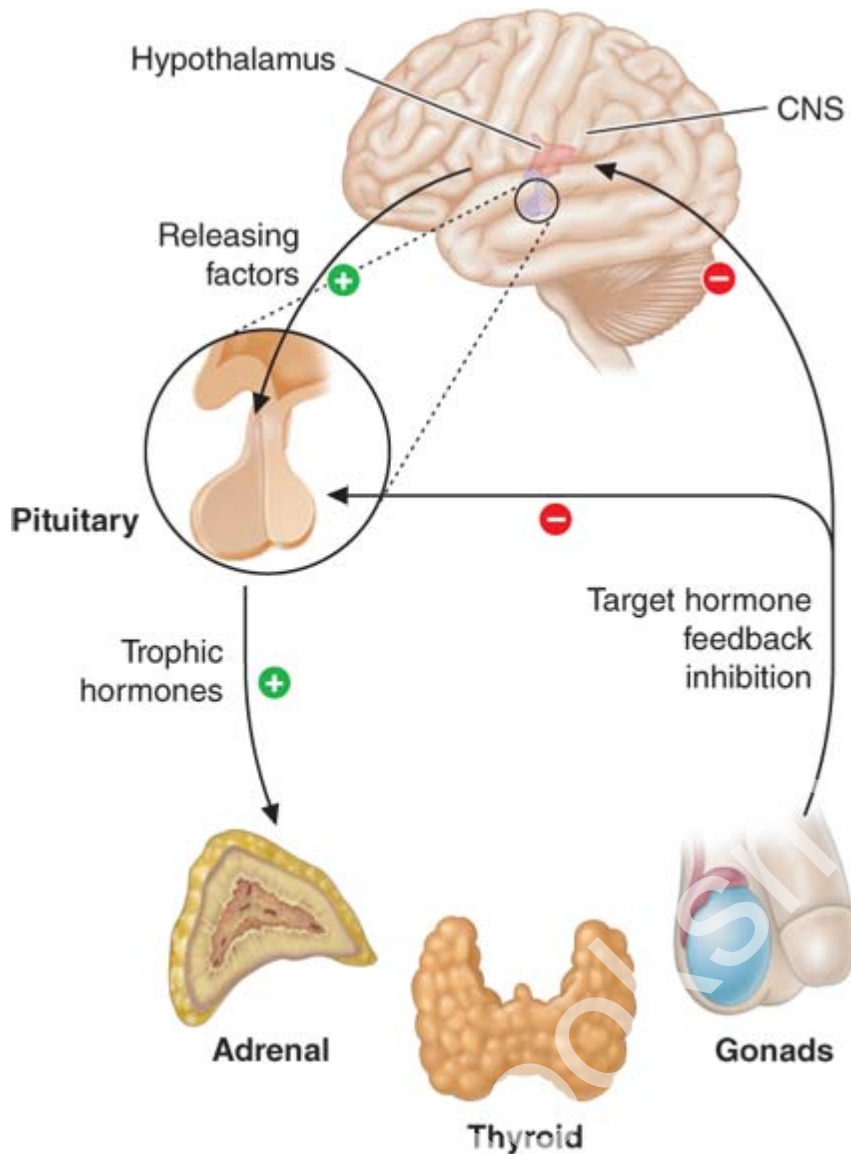
During pregnancy, the increased production of prolactin, in combination with placentally derived steroids (e.g., estrogen and progesterone), prepares the breast for lactation. **Estrogens** induce the production of progesterone receptors, allowing for increased responsiveness to progesterone. In addition to these and other hormones involved in lactation, the nervous system and oxytocin mediate the suckling response and milk release.

HORMONAL FEEDBACK REGULATORY SYSTEMS

Feedback control, both negative and positive, is a fundamental feature of endocrine systems. Each of the major hypothalamic-pituitary-hormone axes is governed by negative feedback, a process that maintains hormone levels within a relatively narrow range (**Chap. 371**). Examples of hypothalamic-pituitary negative feedback include (1) thyroid hormones on the TRH-TSH axis, (2) cortisol on the CRH-ACTH axis, (3) gonadal steroids on the GnRH-LH/FSH axis, and (4) IGF-I on the GHRH-GH axis (**Fig. 370-4**). These regulatory loops include both positive (e.g., TRH, TSH) and negative (e.g., T_4 , T_3) components, allowing for exquisite control of hormone levels. As an example, a small reduction of thyroid hormone triggers a rapid increase of TRH and TSH secretion, resulting in thyroid gland stimulation and increased thyroid hormone production. When thyroid hormone reaches a normal level, it feeds back to suppress TRH and TSH, and a new steady state is attained. Feedback regulation also occurs for endocrine systems that do not involve the pituitary gland, such as calcium feedback on PTH, glucose inhibition of insulin secretion, and leptin feedback on the hypothalamus. An understanding of feedback regulation provides important insights into endocrine testing paradigms (see below).

FIGURE 370-4

Feedback regulation of endocrine axes. CNS, central nervous system.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Positive feedback control also occurs but is not well understood. The primary example is estrogen-mediated stimulation of the midcycle LH surge. Although chronic low levels of estrogen are inhibitory, gradually rising estrogen levels stimulate LH secretion. This effect, which is illustrative of an endocrine rhythm (see below), involves activation of the hypothalamic GnRH pulse generator. In addition, estrogen-primed gonadotropes are extraordinarily sensitive to GnRH, leading to amplification of LH release.

PARACRINE AND AUTOCRINE CONTROL

The previously mentioned examples of feedback control involve classic endocrine pathways in which hormones are released by one gland and act on a distant target gland. However, local regulatory systems, often involving growth factors, are increasingly recognized. *Paracrine regulation* refers to factors released by one cell that act on an adjacent cell in the same tissue. For example, somatostatin secretion by pancreatic islet δ cells inhibits insulin secretion from nearby β cells. *Autocrine regulation* describes the action of a factor

on the same cell from which it is produced. IGF-I acts on many cells that produce it, including chondrocytes, breast epithelium, and gonadal cells. Unlike endocrine actions, paracrine and autocrine control are difficult to document because local growth factor concentrations cannot be measured readily.

Anatomic relationships of glandular systems also greatly influence hormonal exposure: the physical organization of islet cells enhances their intercellular communication; the portal vasculature of the hypothalamic-pituitary system exposes the pituitary to high concentrations of hypothalamic releasing factors; testicular seminiferous tubules gain exposure to high testosterone levels produced by the interdigitated Leydig cells; the pancreas receives nutrient information and local exposure to peptide hormones (incretins) from the gastrointestinal tract; and the liver is the proximal target of insulin action because of portal drainage from the pancreas.

HORMONAL RHYTHMS

The feedback regulatory systems described above are superimposed on hormonal rhythms that are used for adaptation to the environment. Seasonal changes, the daily occurrence of the light-dark cycle, sleep, meals, and stress are examples of the many environmental events that affect hormonal rhythms. The *menstrual cycle* is repeated on average every 28 days, reflecting the time required to follicular maturation and ovulation (**Chap. 385**). Essentially all pituitary hormone rhythms are entrained to sleep and to the *circadian cycle*, generating reproducible patterns that are repeated approximately every 24 h. The HPA axis, for example, exhibits characteristic peaks of ACTH and cortisol production in the early morning, with a nadir during the night. Recognition of these rhythms is important for endocrine testing and treatment. Patients with Cushing's syndrome characteristically exhibit increased midnight cortisol levels compared with normal individuals (**Chap. 379**). In contrast, morning cortisol levels are similar in these groups, as cortisol is normally high at this time of day in normal individuals. The HPA axis is more susceptible to suppression by glucocorticoids administered at night as they blunt the early-morning rise of ACTH. Understanding these rhythms allows glucocorticoid replacement that mimics diurnal production by administering larger doses in the morning than in the afternoon. Disrupted sleep rhythms can alter hormonal regulation. For example, sleep deprivation causes mild insulin resistance, food craving, and hypertension, which are reversible, at least in the short term. Emerging evidence indicates that circadian clock pathways not only regulate sleep-wake cycles but also play important roles in virtually every cell type. For example, tissue-specific deletion of clock genes alters rhythms and levels of gene expression, as well as metabolic responses in liver, adipose, and other tissues.

Other endocrine rhythms occur on a more rapid time scale. Many peptide hormones are secreted in discrete bursts every few hours. LH and FSH secretion are exquisitely sensitive to GnRH pulse frequency. Intermittent pulses of GnRH are required to maintain pituitary sensitivity, whereas continuous exposure to GnRH causes pituitary gonadotrope desensitization. This feature of the hypothalamic-pituitary-gonadotrope axis forms the basis for using long-acting GnRH agonists to treat central precocious puberty or to decrease testosterone levels in the management of prostate cancer. It is important to be aware of the pulsatile nature of hormone secretion and the rhythmic patterns of hormone production in relating serum hormone measurements to

normal values. For some hormones, integrated markers have been developed to circumvent hormonal fluctuations. Examples include 24-h urine collections for cortisol, IGF-I as a biologic marker of GH action, and HbA_{1c} as an index of long-term (weeks to months) blood glucose control.

Often, one must interpret endocrine data only in the context of other hormones. For example, PTH levels typically are assessed in combination with serum calcium concentrations. A high serum calcium level in association with elevated PTH is suggestive of hyperparathyroidism, whereas a suppressed PTH in this situation is more likely to be caused by hypercalcemia of malignancy or other causes of hypercalcemia. Similarly, TSH should be elevated when T₄ and T₃ concentrations are low, reflecting reduced feedback inhibition. When this is not the case, it is important to consider secondary hypothyroidism, which is caused by a defect at the level of the pituitary.

FURTHER READING

Evans RM, Mangelsdorf DJ: Nuclear receptors, RXR, and the big bang. *Cell* 157:255, 2014. [[PubMed: 24679540](#)]

Jameson JL, DeGroot LJ (eds): *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, Elsevier, 2016.

Melmed S et al (eds): *Williams Textbook of Endocrinology*, 13th ed. Philadelphia, Elsevier, 2016.

Moran C, Chatterjee K: Resistance to thyroid hormone α -emerging definition of a disorder of the thyroid hormone action. *J Clin Endocrinol Metab* 101:2317, 2016.

Vassart G, Costagliola S: G protein-coupled receptors: Mutations and endocrine diseases. *Nature Rev Endocrinol* 7:362, 2011.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 371: Physiology of Anterior Pituitary Hormones

Shlomo Melmed; J. Larry Jameson

INTRODUCTION

The anterior pituitary often is referred to as the “master gland” because, together with the hypothalamus, it orchestrates the complex regulatory functions of many other endocrine glands. The anterior pituitary gland produces six major hormones: (1) prolactin (PRL), (2) growth hormone (GH), (3) adrenocorticotrophic hormone (ACTH), (4) luteinizing hormone (LH), (5) follicle-stimulating hormone (FSH), and (6) thyroid-stimulating hormone (TSH) (**Table 371-1**). Pituitary hormones are secreted in a pulsatile manner, reflecting regulation by an array of specific hypothalamic releasing factors. Each of these pituitary hormones elicits specific trophic responses in peripheral target tissues. The hormonal products of those peripheral glands, in turn, exert feedback control at the level of the hypothalamus and pituitary to modulate pituitary function (**Fig. 371-1**). Pituitary tumors cause characteristic hormone excess syndromes. Hormone deficiency may be inherited or acquired. Fortunately, there are efficacious treatments for many pituitary hormone excess and deficiency syndromes. Nonetheless, these diagnoses are often elusive; this emphasizes the importance of recognizing subtle clinical manifestations and performing the correct laboratory diagnostic tests. **For discussion of disorders of the posterior pituitary, or neurohypophysis, see Chap. 374.**

TABLE 371-1

Anterior Pituitary Hormone Expression and Regulation

CELL	CORTICOTROPE	SOMATOTROPE	LACTOTROPE	THYROTROPE	GONADOTROPE
Tissue-specific transcription factor	T-Pit	Prop-1, Pit-1	Prop-1, Pit-1	Prop-1, Pit-1, TEF	SF-1, DAX-1
Fetal appearance	6 weeks	8 weeks	12 weeks	12 weeks	12 weeks
Hormone	POMC	GH	PRL	TSH	FSH, LH
Protein	Polypeptide	Polypeptide	Polypeptide	Glycoprotein α , β subunits	Glycoprotein α , β subunits
Amino acids	266 (ACTH 1–39)	191	198	211	210, 204
Stimulators	CRH, AVP, gp-130 cytokines	GHRH, ghrelin	Estrogen, TRH, VIP	TRH	GnRH, activins, estrogen
Inhibitors	Glucocorticoids	Somatostatin, IGF-I	Dopamine	T ₃ , T ₄ , dopamine, somatostatin, glucocorticoids	Sex steroids, inhibin
Target gland	Adrenal	Liver, bone, other tissues	Breast, other tissues	Thyroid	Ovary, testis
Trophic effect	Steroid production	IGF-I production, growth induction, insulin antagonism	Milk production	T ₄ synthesis and secretion	Sex steroid production, follicle growth, germ cell maturation

CELL	CORTICOTROPE	SOMATOTROPE	LACTOTROPE	THYROTROPE	GONADOTROPE
Normal range	ACTH, 4–22 pg/L	<0.5 µg/L ^a	M <15 µg/L; F <20 µg/L	0.1–5 mU/L	M, 5–20 IU/L; F (basal), 5–20 IU/L

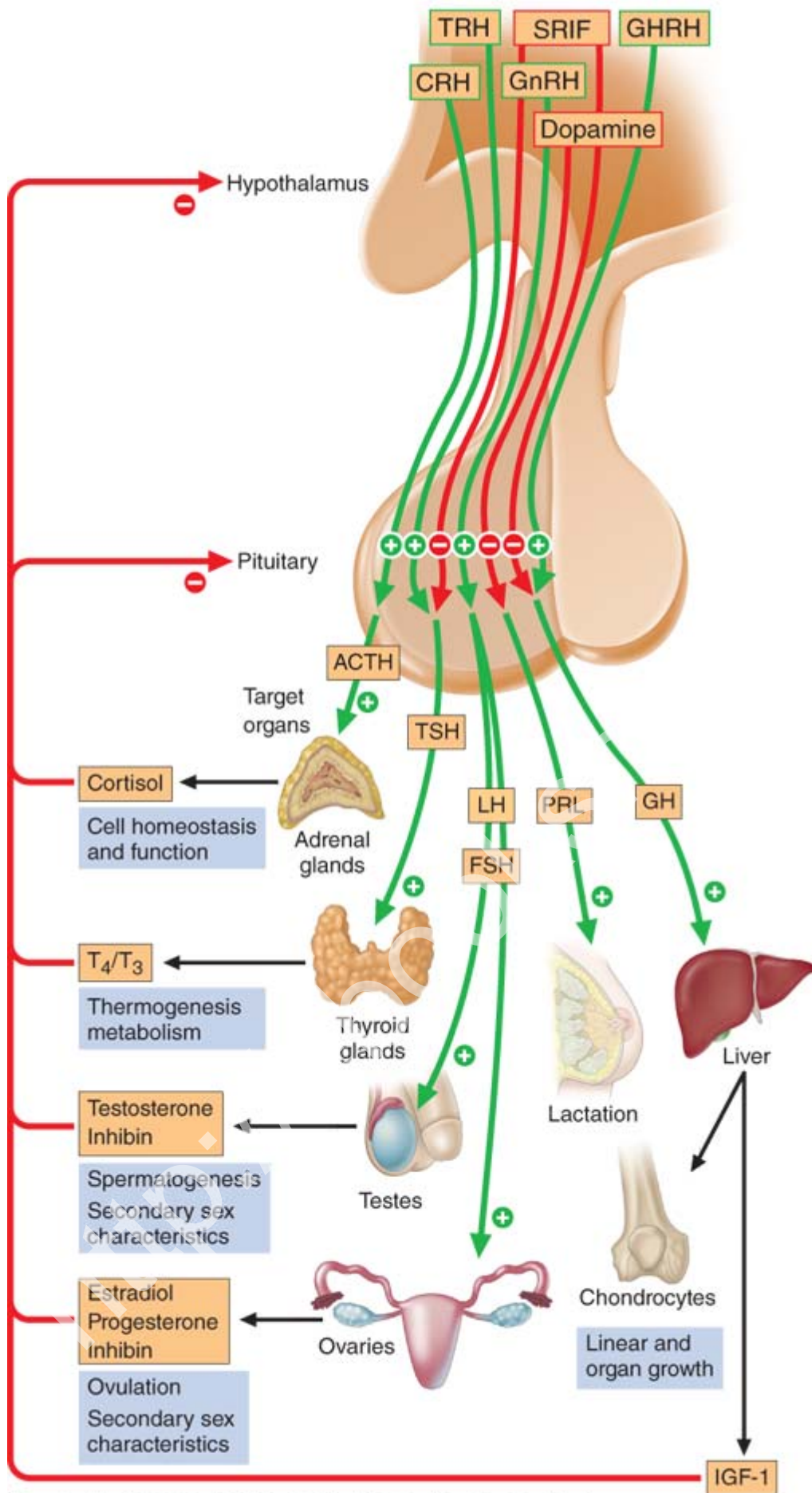
^aHormone secretion integrated over 24 h.

Abbreviations: M, male; F, female. For other abbreviations, see text.

Source: Adapted from S Melmed: Hypothalamic-pituitary regulation, in P Conn (ed): Conn's Translational Neuroscience. San Diego, CA, Elsevier, 2017.

FIGURE 371-1

Diagram of pituitary axes. Hypothalamic hormones regulate anterior pituitary trophic hormones that in turn determine target gland secretion. Peripheral hormones feed back to regulate hypothalamic and pituitary hormones. For abbreviations, see text.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

ANATOMY AND DEVELOPMENT

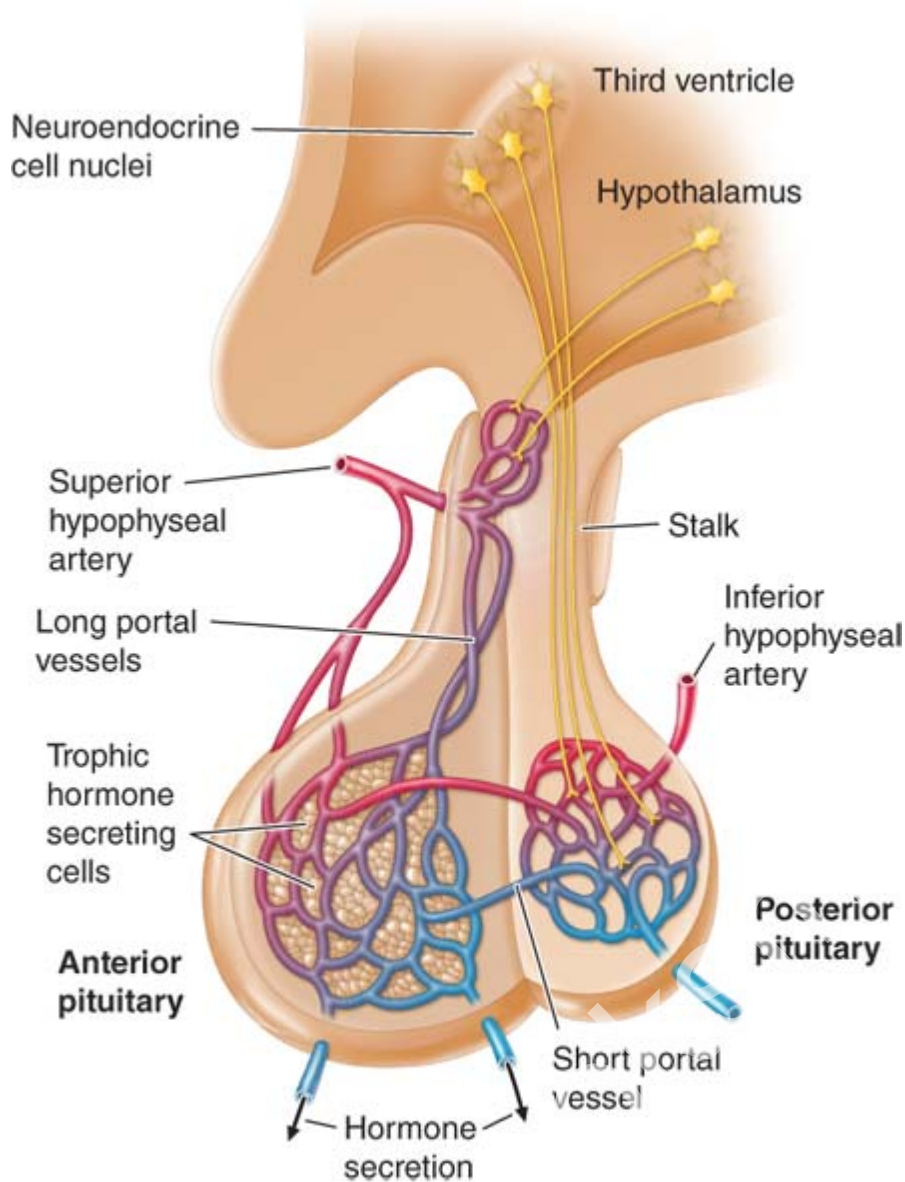
ANATOMY

The pituitary gland weighs ~600 mg and is located within the sella turcica ventral to the diaphragma sella; it consists of anatomically and functionally distinct anterior and posterior lobes. The bony sella is contiguous to vascular and neurologic structures, including the cavernous sinuses, cranial nerves, and optic chiasm. Thus, expanding intrasellar pathologic processes may have significant central mass effects in addition to their endocrinologic impact.

Hypothalamic neural cells synthesize specific releasing and inhibiting hormones that are secreted directly into the portal vessels of the pituitary stalk. Blood supply of the pituitary gland comes from the superior and inferior hypophyseal arteries (Fig. 371-2). The hypothalamic-pituitary portal plexus provides the major blood source for the anterior pituitary, allowing reliable transmission of hypothalamic peptide pulses without significant systemic dilution; consequently, pituitary cells are exposed to releasing or inhibiting factors and in turn release their respective hormones as discrete pulses into the systemic circulation (Fig. 371-3).

FIGURE 371-2

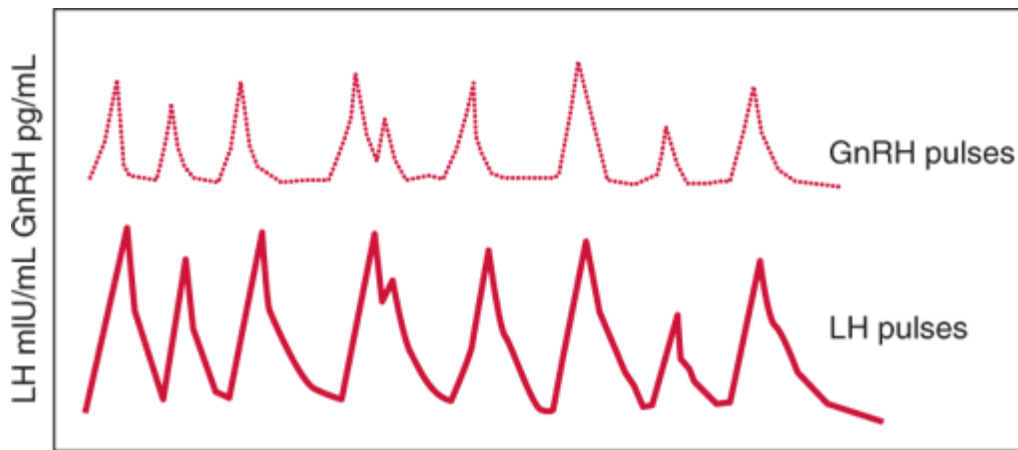
Diagram of hypothalamic-pituitary vasculature. The hypothalamic nuclei produce hormones that traverse the portal system and impinge on anterior pituitary cells to regulate pituitary hormone secretion. Posterior pituitary hormones are derived from direct neural extensions.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE 371-3

Hypothalamic gonadotropin-releasing hormone (GnRH) pulses induce secretory pulses of luteinizing hormone (LH).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The posterior pituitary is supplied by the inferior hypophyseal arteries. In contrast to the anterior pituitary, the posterior lobe is directly innervated by hypothalamic neurons (supraopticohypophyseal and tuberohypophyseal nerve tracts) via the pituitary stalk (**Chap. 374**). Thus, posterior pituitary production of vasopressin (antidiuretic hormone [ADH]) and oxytocin is particularly sensitive to neuronal damage by lesions that affect the pituitary stalk or hypothalamus.

PITUITARY DEVELOPMENT

The embryonic differentiation and maturation of anterior pituitary cells have been elucidated in considerable detail. Pituitary development from Rathke's pouch involves a complex interplay of lineage-specific transcription factors expressed in pluripotent precursor cells and gradients of locally produced growth factors (**Table 371-1**). The transcription factor Prop-1 induces pituitary development of Pit-1-specific lineages as well as gonadotropes. The transcription factor Pit-1 determines cell-specific expression of GH, PRL, and TSH in somatotropes, lactotropes, and thyrotropes. Expression of high levels of estrogen receptors in cells that contain Pit-1 favors PRL expression, whereas thyrotrope embryonic factor (TEF) induces TSH expression. Pit-1 binds to GH, PRL, and TSH gene regulatory elements as well as to recognition sites on its own promoter, providing a mechanism for determining specific pituitary hormone phenotypic stability. Gonadotrope cell development is further defined by the cell-specific expression of the nuclear receptors steroidogenic factor (SF-1) and dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX-1). Development of corticotrope cells, which express the proopiomelanocortin (*POMC*) gene, requires the T-Pit transcription factor. Abnormalities of pituitary development caused by mutations of Pit-1, Prop-1, SF-1, DAX-1, and T-Pit result in rare, selective or combined pituitary hormone deficit syndromes.

ANTERIOR PITUITARY HORMONES

Each anterior pituitary hormone is under unique control, and each exhibits highly specific normal and dysregulated secretory characteristics.

PROLACTIN

Synthesis

PRL consists of 198 amino acids and has a molecular mass of 21,500 kDa; it is weakly homologous to GH and human placental lactogen (hPL), reflecting the duplication and divergence of a common GH-PRL-hPL precursor gene. PRL is synthesized in lactotrobes, which constitute about 20% of anterior pituitary cells. Lactotrobes and somatotrobes are derived from a common precursor cell that may give rise to a tumor that secretes both PRL and GH. Marked lactotrope cell hyperplasia develops during pregnancy and the first few months of lactation. These transient functional changes in the lactotrope population are induced by estrogen.

Secretion

Normal adult serum PRL levels are about 10–25 µg/L in women and 10–20 µg/L in men. PRL secretion is pulsatile, with the highest secretory peaks occurring during rapid eye movement sleep. Peak serum PRL levels (up to 30 µg/L) occur between 4:00 and 6:00 A.M. The circulating half-life of PRL is about 50 min.

PRL is unique among the pituitary hormones in that the predominant central control mechanism is inhibitory, reflecting tonic dopamine-mediated suppression of PRL release. This regulatory pathway accounts for the spontaneous PRL hypersecretion that occurs with pituitary stalk section, often a consequence of head trauma or compressive mass lesions at the skull base. Pituitary dopamine type 2 (D₂) receptors mediate inhibition of PRL synthesis and secretion. Targeted disruption (gene knockout) of the murine D₂ receptor in mice results in hyperprolactinemia and lactotrope proliferation. As discussed below, dopamine agonists play a central role in the management of hyperprolactinemic disorders.

Thyrotropin-releasing hormone (TRH) (pyro Glu-His-Pro-NH₂) is a hypothalamic tripeptide that elicits PRL release within 15–30 min after intravenous injection. TRH primarily regulates TSH, and the physiologic relevance of TRH for PRL regulation is unclear ([Chap. 375](#)). *Vasoactive intestinal peptide* (VIP) also induces PRL release, whereas glucocorticoids and thyroid hormone weakly suppress PRL secretion.

Serum PRL levels rise transiently after exercise, meals, sexual intercourse, minor surgical procedures, general anesthesia, chest wall injury, acute myocardial infarction, and other forms of acute stress. PRL levels increase markedly (about tenfold) during pregnancy and decline rapidly within 2 weeks of parturition. If breast-feeding is initiated, basal PRL levels remain elevated; suckling stimulates transient reflex increases in PRL levels that last for about 30–45 min. Breast suckling activates afferent neural pathways in the hypothalamus that induce PRL release. With time, suckling-induced responses diminish and interfeeding PRL levels return to normal.

Action

The PRL receptor is a member of the type I cytokine receptor family that also includes GH and interleukin (IL) 6 receptors. Ligand binding induces receptor dimerization and intracellular signaling by Janus kinase (JAK), which stimulates translocation of the signal transduction and activators of transcription (STAT) family to

activate target genes. Heterozygous mutations of the PRL receptor result in PRL insensitivity, hyperprolactinemia, and oligomenorrhea. In the breast, the lobuloalveolar epithelium proliferates in response to PRL, placental lactogens, estrogen, progesterone, and local paracrine growth factors, including insulin-like growth factor I (IGF-I).

PRL acts to induce and maintain lactation, decrease reproductive function, and suppress sexual drive. These functions are geared toward ensuring that maternal lactation is sustained and not interrupted by pregnancy. PRL inhibits reproductive function by suppressing hypothalamic gonadotropin-releasing hormone (GnRH) and pituitary gonadotropin secretion and by impairing gonadal steroidogenesis in both women and men. In the ovary, PRL blocks folliculogenesis and inhibits granulosa cell aromatase activity, leading to hypoestrogenism and anovulation. PRL also has a luteolytic effect, generating a shortened, or inadequate, luteal phase of the menstrual cycle. In men, attenuated LH secretion leads to low testosterone levels and decreased spermatogenesis. These hormonal changes decrease libido and reduce fertility in patients with hyperprolactinemia.

GROWTH HORMONE

Synthesis

GH is the most abundant anterior pituitary hormone, and GH-secreting somatotrope cells constitute up to 50% of the total anterior pituitary cell population. Mammosomatotrope cells, which coexpress PRL with GH, can be identified by using double immunostaining techniques. Somatotrope development and GH transcription are determined by expression of the cell-specific Pit-1 nuclear transcription factor. Five distinct genes encode GH and related proteins. The pituitary GH gene (*hGH-M*) produces two alternatively spliced products that give rise to 22-kDa GH (191 amino acids) and a less abundant 20-kDa GH molecule with similar biologic activity. Placental syncytiotrophoblast cells express a GH variant (*hGH-V*) gene; the related hormone human chorionic somatotropin (HCS) is expressed by distinct members of the gene cluster.

Secretion

GH secretion is controlled by complex hypothalamic and peripheral factors. *GH-releasing hormone* (GHRH) is a 44-amino-acid hypothalamic peptide that stimulates GH synthesis and release. Ghrelin, an octanoylated gastric-derived peptide, and synthetic agonists of the *GHS-R* induce GHRH and also directly stimulate GH release. *Somatostatin* (somatotropin-release inhibiting factor [SRIF]) is synthesized in the medial preoptic area of the hypothalamus and inhibits GH secretion. GHRH is secreted in discrete spikes that elicit GH pulses, whereas SRIF sets basal GH secretory tone. SRIF also is expressed in many extrahypothalamic tissues, including the central nervous system (CNS), gastrointestinal tract, and pancreas, where it also acts to inhibit islet hormone secretion. *IGF-I*, the peripheral target hormone for GH, feeds back to inhibit GH; estrogen induces GH, whereas chronic glucocorticoid excess suppresses GH release.

Surface receptors on the somatotrope regulate GH synthesis and secretion. The GHRH receptor is a G protein-coupled receptor (GPCR) that signals through the intracellular cyclic AMP pathway to stimulate

somatotrope cell proliferation as well as GH production. Inactivating mutations of the GHRH receptor cause profound dwarfism. A distinct surface receptor for ghrelin, the gastric-derived GH secretagogue, is expressed in both the hypothalamus and pituitary. Somatostatin binds to five distinct receptor subtypes (SST1 to SST5); SST2 and SST5 subtypes preferentially suppress GH (and TSH) secretion and SST5 signals to suppress ACTH secretion.

GH secretion is pulsatile, with highest peak levels occurring at night, generally correlating with sleep onset. GH secretory rates decline markedly with age so that hormone levels in middle age are about 15% of pubertal levels. These changes are paralleled by an age-related decline in lean muscle mass. GH secretion is also reduced in obese individuals, although IGF-I levels may not be suppressed, suggesting a change in the setpoint for feedback control. Elevated GH levels occur within an hour of deep sleep onset as well as after exercise, physical stress, and trauma and during sepsis. Integrated 24-h GH secretion is higher in women and is also enhanced by estrogen replacement, likely reflective of increased peripheral GH resistance. Using standard assays, random GH measurements are undetectable in ~50% of daytime samples obtained from healthy subjects and are also undetectable in most obese and elderly subjects. Thus, single random GH measurements do not distinguish patients with adult GH deficiency from normal persons.

GH secretion is profoundly influenced by nutritional factors. Using newer ultrasensitive GH assays with a sensitivity of 0.002 µg/L, a glucose load suppresses GH to <0.7 µg/L in women and to <0.07 µg/L in men. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition or prolonged fasting. GH is stimulated by intravenous L-arginine, dopamine, and apomorphine (a dopamine receptor agonist), as well as by α -adrenergic pathways. β -Adrenergic blockade induces basal GH and enhances GHRH- and insulin-evoked GH release.

Action

The pattern of GH secretion may affect tissue responses. The higher GH pulsatility observed in men compared with the relatively continuous basal GH secretion in women may be an important biologic determinant of linear growth patterns and liver enzyme induction.

The 70-kDa peripheral GH receptor protein has structural homology with the cytokine/hematopoietic superfamily. A fragment of the receptor extracellular domain generates a soluble GH binding protein (GHBP) that interacts with GH in the circulation. The liver and cartilage express the greatest number of GH receptors. GH binding to preformed receptor dimers is followed by internal rotation and subsequent signaling through the JAK/STAT pathway. Activated STAT proteins translocate to the nucleus, where they modulate expression of GH-regulated target genes. GH analogues that bind to the receptor but are incapable of mediating receptor signaling are potent antagonists of GH action. A GH receptor antagonist (pegvisomant) is approved for treatment of acromegaly.

GH induces protein synthesis and nitrogen retention and also impairs glucose tolerance by antagonizing insulin action. GH also stimulates lipolysis, leading to increased circulating fatty acid levels, reduced omental fat mass, and enhanced lean body mass. GH promotes sodium, potassium, and water retention and elevates

serum levels of inorganic phosphate. Linear bone growth occurs as a result of complex hormonal and growth factor actions, including those of IGF-I. GH stimulates epiphyseal prechondrocyte differentiation. These precursor cells produce IGF-I locally, and their proliferation is also responsive to the growth factor.

Insulin-Like Growth Factors

Although GH exerts direct effects in target tissues, many of its physiologic effects are mediated indirectly through IGF-I, a potent growth and differentiation factor. The liver is the major source of circulating IGF-I. In peripheral tissues, IGF-I also exerts local paracrine actions that appear to be both dependent on and independent of GH. Thus, GH administration induces circulating IGF-I as well as stimulating local IGF-I production in multiple tissues.

Both IGF-I and IGF-II are bound to high-affinity circulating IGF-binding proteins (IGFBPs) that regulate IGF availability and bioactivity. Levels of IGFBP3 are GH-dependent, and it serves as the major carrier protein for circulating IGF-I. GH deficiency and malnutrition usually are associated with low IGFBP3 levels. IGFBP1 and IGFBP2 regulate local tissue IGF action but do not bind appreciable amounts of circulating IGF-I.

Serum IGF-I concentrations are profoundly affected by physiologic factors. Levels increase during puberty, peak at 16 years, and subsequently decline by >80% during the aging process. IGF-I concentrations are higher in women than in men. Because GH is the major determinant of hepatic IGF-I synthesis, abnormalities of GH synthesis or action (e.g., pituitary failure, GHRH receptor defect, GH receptor defect or pharmacologic GH receptor blockade) reduce IGF-I levels. Hypocaloric states are associated with GH resistance; IGF-I levels are therefore low with cachexia, malnutrition, and sepsis. In acromegaly, IGF-I levels are invariably high and reflect a log-linear relationship with circulating GH concentrations.

IGF-I PHYSIOLOGY

Injected IGF-I (100 µg/kg) induces hypoglycemia, and lower doses improve insulin sensitivity in patients with severe insulin resistance and diabetes. In cachectic subjects, IGF-I infusion (12 µg/kg per h) enhances nitrogen retention and lowers cholesterol levels. Longer-term subcutaneous IGF-I injections enhance protein synthesis and are anabolic. Although bone formation markers are induced, bone turnover also may be stimulated by IGF-I. IGF-I is approved for use in patients with GH-resistance syndromes.

IGF-I side effects are dose-dependent, and overdose may result in hypoglycemia, hypotension, fluid retention, temporomandibular jaw pain, and increased intracranial pressure, all of which are reversible. Avascular femoral head necrosis has been reported. Chronic excess IGF-I administration presumably would result in features of acromegaly.

ADRENOCORTICOTROPIC HORMONE

(See also [Chap. 379](#))

Synthesis

ACTH-secreting corticotrope cells constitute about 20% of the pituitary cell population. ACTH (39 amino acids) is derived from the POMC precursor protein (266 amino acids) that also generates several other peptides, including β -lipotropin, β -endorphin, met-enkephalin, α -melanocyte-stimulating hormone (α -MSH), and corticotropin-like intermediate lobe protein (CLIP). The *POMC* gene is potently suppressed by glucocorticoids and induced by corticotropin-releasing hormone (CRH), arginine vasopressin (AVP), and proinflammatory cytokines, including IL-6, as well as leukemia inhibitory factor.

CRH, a 41-amino-acid hypothalamic peptide synthesized in the paraventricular nucleus as well as in higher brain centers, is the predominant stimulator of ACTH synthesis and release. The CRH receptor is a GPCR that is expressed on the corticotrope and signals to induce *POMC* transcription.

Secretion

ACTH secretion is pulsatile and exhibits a characteristic circadian rhythm, peaking at about 6:00 A.M. and reaching a nadir about midnight. Adrenal glucocorticoid secretion, which is driven by ACTH, follows a parallel diurnal pattern. ACTH circadian rhythmicity is determined by variations in secretory pulse amplitude rather than changes in pulse frequency. Superimposed on this endogenous rhythm, ACTH levels are increased by physical and psychological stress, exercise, acute illness, and insulin-induced hypoglycemia.

Glucocorticoid-mediated negative regulation of the hypothalamic-pituitary-adrenal (HPA) axis occurs as a consequence of both hypothalamic CRH suppression and direct attenuation of pituitary *POMC* gene expression and ACTH release. In contrast, loss of cortisol feedback inhibition, as occurs in primary adrenal failure, results in extremely high ACTH levels.

Acute inflammatory or septic insults activate the HPA axis through the integrated actions of proinflammatory cytokines, bacterial toxins, and neural signals. The overlapping cascade of ACTH-inducing cytokines (tumor necrosis factor [TNF]; IL-1, -2, and -6; and leukemia inhibitory factor) activates hypothalamic CRH and AVP secretion, pituitary *POMC* gene expression, and local pituitary paracrine cytokine networks. The resulting cortisol elevation restrains the inflammatory response and enables host protection. Concomitantly, cytokine-mediated central glucocorticoid receptor resistance impairs glucocorticoid suppression of the HPA. Thus, the neuroendocrine stress response reflects the net result of highly integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals acting to regulate cortisol secretion.

Action

The major function of the HPA axis is to maintain metabolic homeostasis and mediate the neuroendocrine stress response. ACTH induces adrenocortical steroidogenesis by sustaining adrenal cell proliferation and function. The receptor for ACTH, designated *melanocortin-2 receptor*, is a GPCR that induces steroidogenesis by stimulating a cascade of steroidogenic enzymes (**Chap. 379**).

GONADOTROPINS: FSH AND LH

Synthesis and Secretion

Gonadotrope cells constitute about 10% of anterior pituitary cells and produce two gonadotropin hormones —LH and FSH. Like TSH and hCG, LH and FSH are glycoprotein hormones that comprise α and β subunits. The α subunit is common to these glycoprotein hormones; specificity of hormone function is conferred by the β subunits, which are expressed by separate genes.

Gonadotropin synthesis and release are dynamically regulated. This is particularly true in women, in whom rapidly fluctuating gonadal steroid levels vary throughout the menstrual cycle. Hypothalamic GnRH, a 10-amino-acid peptide, regulates the synthesis and secretion of both LH and FSH. Brain kisspeptin, a product of the *KISS1* gene, regulates hypothalamic GnRH release. GnRH is secreted in discrete pulses every 60–120 min, and the pulses in turn elicit LH and FSH pulses (Fig. 371-3). The pulsatile mode of GnRH input is essential to its action; pulses prime gonadotrope responsiveness, whereas continuous GnRH exposure induces desensitization. Based on this phenomenon, long-acting GnRH agonists are used to suppress gonadotropin levels in children with precocious puberty and in men with prostate cancer (Chap. 83) and are used in some ovulation-induction protocols to reduce levels of endogenous gonadotropins (Chap. 385). Estrogens act at both the hypothalamus and the pituitary to modulate gonadotropin secretion. Chronic estrogen exposure is inhibitory, whereas rising estrogen levels, as occur during the preovulatory surge, exert positive feedback to increase gonadotropin pulse frequency and amplitude. Progesterone slows GnRH pulse frequency but enhances gonadotropin responses to GnRH. Testosterone feedback in men also occurs at the hypothalamic and pituitary levels and is mediated in part by its conversion to estrogens.

Although GnRH is the main regulator of LH and FSH secretion, FSH synthesis is also under separate control by the gonadal peptides inhibin and activin, which are members of the transforming growth factor β (TGF- β) family. Inhibin selectively suppresses FSH, whereas activin stimulates FSH synthesis (Chap. 385).

Action

The gonadotropin hormones interact with their respective GPCRs expressed in the ovary and testis, evoking germ cell development and maturation and steroid hormone biosynthesis. In women, FSH regulates ovarian follicle development and stimulates ovarian estrogen production. LH mediates ovulation and maintenance of the corpus luteum. In men, LH induces Leydig cell testosterone synthesis and secretion, and FSH stimulates seminiferous tubule development and regulates spermatogenesis.

THYROID-STIMULATING HORMONE

Synthesis and Secretion

TSH-secreting thyrotrope cells constitute 5% of the anterior pituitary cell population. TSH shares a common α subunit with LH and FSH but contains a specific TSH β subunit. TRH is a hypothalamic tripeptide (pyroglutamyl histidylprolinamide) that acts through a pituitary GPCR to stimulate TSH synthesis and secretion; it also stimulates the lactotrope cell to secrete PRL. TSH secretion is stimulated by TRH, whereas thyroid hormones, dopamine, somatostatin, and glucocorticoids suppress TSH by overriding TRH induction.

Thyrotrope cell proliferation and TSH secretion are both induced when negative feedback inhibition by thyroid hormones is removed. Thus, thyroid damage (including surgical thyroidectomy), radiation-induced hypothyroidism, chronic thyroiditis, and prolonged goitrogen exposure are associated with increased TSH levels. Long-standing untreated hypothyroidism can lead to elevated TSH levels, which may be associated with thyrotrope hyperplasia and pituitary enlargement and may sometimes be evident on magnetic resonance imaging.

Action

TSH is secreted in pulses, although the excursions are modest in comparison to other pituitary hormones because of the low amplitude of the pulses and the relatively long half-life of TSH. Consequently, single determinations of TSH suffice to precisely assess its circulating levels. TSH binds to a GPCR on thyroid follicular cells to stimulate thyroid hormone synthesis and release (**Chap. 375**).

FURTHER READING

Bernard V et al: New insights in prolactin: Pathological implications. *Nat Rev Endocrinol* 11:265, 2015. [[PubMed: 25781857](#)]

Brooks AJ, Waters MJ: The growth hormone receptor: Mechanism of activation and clinical implications. *Nat Rev Endocrinol* 6:515, 2010. [[PubMed: 20664532](#)]

Ho K: The year in pituitary 2014. *J Clin Endocrinol Metab* 99:4449, 2014. [[PubMed: 25322276](#)]

Kelberman D et al: Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev* 30:790, 2009. [[PubMed: 19837867](#)]

Langlais D et al: Adult pituitary cell maintenance: Lineage-specific contribution of self-duplication. *Mol Endocrinol* 27:1103, 2013. [[PubMed: 23754801](#)]

Murray PG, Higham CE, Clayton PE: 60 years of neuroendocrinology: The hypothalamo-GH axis: The past 60 years. *J Endocrinol* 226:T123, 2015. [[PubMed: 26040485](#)]

Newey PJ et al: Mutant prolactin receptor and familial hyperprolactinemia. *N Engl J Med* 369:2012, 2013.

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

Silverchair

Chapter 372: Hypopituitarism

Shlomo Melmed; J. Larry Jameson

FIGURE 372-1

INTRODUCTION

Inadequate production of anterior pituitary hormones leads to features of hypopituitarism. Impaired production of one or more of the anterior pituitary trophic hormones can result from inherited disorders; more commonly, adult hypopituitarism is acquired and reflects the compressive mass effects of tumors or the consequences of local pituitary or hypothalamic traumatic, inflammatory, or vascular damage. These processes also may impair synthesis or secretion of hypothalamic hormones, with resultant pituitary failure ([Table 372-1](#)).

TABLE 372-1

Etiology of Hypopituitarism^a

Development/structural

Transcription factor defect

Pituitary dysplasia/aplasia

Congenital central nervous system mass, encephalocele

Primary empty sella

Congenital hypothalamic disorders (septo-optic dysplasia, Prader-Willi syndrome, Bardet-Biedl syndrome, Kallmann syndrome)

Traumatic

Surgical resection

Radiation damage

Head injuries

Neoplastic

Pituitary adenoma

Parasellar mass (germinoma, ependymoma, glioma)

Rathke's cyst

Craniopharyngioma

Hypothalamic hamartoma, gangliocytoma

Pituitary metastases (breast, lung, colon carcinoma)

Lymphoma and leukemia

Meningioma

Infiltrative/inflammatory

Lymphocytic hypophysitis

Hemochromatosis

Sarcoidosis

Histiocytosis X

Granulomatous hypophysitis

Transcription factor antibodies

Immunotherapy

Vascular

Pituitary apoplexy

Pregnancy-related (infarction with diabetes; postpartum necrosis)

Sickle cell disease

Arteritis

Infections

Fungal (histoplasmosis)

Parasitic (toxoplasmosis)

Tuberculosis

^aTrophic hormone failure associated with pituitary compression or destruction usually occurs sequentially: growth hormone > follicle-stimulating hormone > luteinizing hormone > thyroid-stimulating hormone > adrenocorticotrophic hormone. During childhood, growth retardation is often the presenting feature, and in adults, hypogonadism is the earliest symptom.

DEVELOPMENTAL AND GENETIC CAUSES OF HYPOPITUITARISM

Pituitary Dysplasia

Pituitary dysplasia may result in aplastic, hypoplastic, or ectopic pituitary gland development. Because pituitary development follows midline cell migration from the nasopharyngeal Rathke's pouch, midline craniofacial disorders may be associated with pituitary dysplasia. Acquired pituitary failure in the newborn also can be caused by birth trauma, including cranial hemorrhage, asphyxia, and breech delivery.

SEPTO-OPTIC DYSPLASIA

Hypothalamic dysfunction and hypopituitarism may result from dysgenesis of the septum pellucidum or corpus callosum. Affected children have mutations in the *HESX1* gene, which is involved in early development of the ventral prosencephalon. These children exhibit variable combinations of cleft palate, syndactyly, ear deformities, hypertelorism, optic nerve hypoplasia, micropenis, and anosmia. Pituitary dysfunction leads to diabetes insipidus, growth hormone (GH) deficiency and short stature, and, occasionally, thyroid-stimulating hormone (TSH) deficiency.

Tissue-Specific Factor Mutations

Several pituitary cell-specific transcription factors, such as Pit-1 and Prop-1, are critical for determining the development and committed function of differentiated anterior pituitary cell lineages. Autosomal dominant or recessive Pit-1 mutations cause combined GH, prolactin (PRL), and TSH deficiencies. These patients usually present with growth failure and varying degrees of hypothyroidism. The pituitary may appear hypoplastic on magnetic resonance imaging (MRI).

Prop-1 is expressed early in pituitary development and appears to be required for Pit-1 function. Familial and sporadic *PROP1* mutations result in combined GH, PRL, TSH, and gonadotropin deficiency. Over 80% of these patients have growth retardation; by adulthood, all are deficient in TSH and gonadotropins, and a small minority later develop adrenocorticotrophic hormone (ACTH) deficiency. Because of gonadotropin deficiency, these individuals do not enter puberty spontaneously. In some cases, the pituitary gland appears enlarged on MRI. *TPIT* mutations result in ACTH deficiency associated with hypocortisolism.

Developmental Hypothalamic Dysfunction

KALLMANN SYNDROME

Kallmann syndrome results from defective hypothalamic gonadotropin-releasing hormone (GnRH) synthesis and is associated with anosmia or hyposmia due to olfactory bulb agenesis or hypoplasia ([Chap. 384](#)). Classically, the syndrome may also be associated with color blindness, optic atrophy, nerve deafness, cleft palate, renal abnormalities, cryptorchidism, and neurologic abnormalities such as mirror movements. The initial genetic cause was identified in the X-linked *KAL* gene, mutations of which impair embryonic migration of GnRH neurons from the hypothalamic olfactory placode to the hypothalamus. Since then, at least a dozen additional genetic abnormalities, in addition to *KAL* mutations, have been found to cause isolated GnRH deficiency. Autosomal recessive (i.e., *GPR54*, *KISS1*) and dominant (i.e., *FGFR1*) modes of transmission have been described, and there is a growing list of genes associated with GnRH deficiency (including *GNRH1*, *PROK2*, *PROKR2*, *CHD7*, *PCSK1*, *FGF8*, *NELF*, *WDR11*, *TAC3*, *TACR3*, and *SEMA3E*). Some patients have oligogenic mutations. Associated clinical features, in addition to GnRH deficiency, vary depending on the genetic cause. GnRH deficiency prevents progression through puberty. Males present with delayed puberty and pronounced hypogonadal features, including micropenis, probably the result of low testosterone levels during infancy. Females present with primary amenorrhea and failure of secondary sexual development.

Kallmann syndrome and other causes of congenital GnRH deficiency are characterized by low luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels and low concentrations of sex steroids (testosterone or [estradiol](#)). In sporadic cases of isolated gonadotropin deficiency, the diagnosis is often one of exclusion after other known causes of hypothalamic-pituitary dysfunction have been eliminated. Repetitive GnRH administration restores normal pituitary gonadotropin responses, pointing to a hypothalamic defect in these patients.

Long-term treatment of males with human chorionic gonadotropin (hCG) or testosterone restores pubertal development and secondary sex characteristics; women can be treated with cyclic estrogen and progestin. Fertility may be restored by the administration of gonadotropins or by using a portable infusion pump to deliver subcutaneous, pulsatile GnRH.

BARDET-BIEDL SYNDROME

This very rare genetically heterogeneous disorder is characterized by mental retardation, renal abnormalities, obesity, and hexadactyly, brachydactyly, or syndactyly. Central diabetes insipidus may or may not be associated. GnRH deficiency occurs in 75% of males and half of affected females. Retinal degeneration begins in early childhood, and most patients are blind by age 30. Numerous subtypes of Bardet-Biedl syndrome (BBS) have been identified, with genetic linkage to at least nine different loci. Several of the loci encode genes involved in basal body cilia function, and this may account for the diverse clinical manifestations.

LEPTIN AND LEPTIN RECEPTOR MUTATIONS

Deficiencies of leptin or its receptor cause a broad spectrum of hypothalamic abnormalities, including hyperphagia, obesity, and central hypogonadism ([Chap. 394](#)). Decreased GnRH production in these patients results in attenuated pituitary FSH and LH synthesis and release.

PRADER-WILLI SYNDROME

This is a contiguous gene syndrome that results from deletion of the paternal copies of the imprinted *SNRPN* gene, the *NECDIN* gene, and possibly other genes on chromosome 15q. Prader-Willi syndrome is associated with hypogonadotropic hypogonadism, hyperphagia-obesity, chronic muscle hypotonia, mental retardation, and adult-onset diabetes mellitus. Multiple somatic defects also involve the skull, eyes, ears, hands, and feet. Diminished hypothalamic oxytocin- and vasopressin-producing nuclei have been reported. Deficient GnRH synthesis is suggested by the observation that chronic GnRH treatment restores pituitary LH and FSH release.

ACQUIRED HYPOPITUITARISM

Hypopituitarism may be caused by accidental or neurosurgical trauma; vascular events such as apoplexy; pituitary or hypothalamic neoplasms, craniopharyngioma, lymphoma, or metastatic tumors; inflammatory disease such as lymphocytic hypophysitis; autoimmune hypophysitis associated with checkpoint inhibitor cancer immunotherapy; infiltrative disorders such as sarcoidosis, hemochromatosis ([Chap. 407](#)), and tuberculosis; or irradiation.

Increasing evidence suggests that patients with brain injury, including contact sports trauma, subarachnoid hemorrhage, and irradiation, have transient hypopituitarism and require intermittent long-term endocrine follow-up, because permanent hypothalamic or pituitary dysfunction will develop in 25–40% of these patients.

Hypothalamic Infiltration Disorders

These disorders—including sarcoidosis, histiocytosis X, amyloidosis, and hemochromatosis—frequently involve both hypothalamic and pituitary neuronal and neurochemical tracts. Consequently, diabetes insipidus occurs in half of patients with these disorders. Growth retardation is seen if attenuated GH secretion occurs before puberty. Hypogonadotropic hypogonadism and hyperprolactinemia are also common.

Inflammatory Lesions

Pituitary damage and subsequent secretory dysfunction can be seen with chronic site infections such as tuberculosis, with opportunistic fungal infections associated with AIDS, and in tertiary syphilis. Other inflammatory processes, such as granulomas and sarcoidosis, may mimic the features of a pituitary adenoma. These lesions may cause extensive hypothalamic and pituitary damage, leading to trophic hormone deficiencies.

Cranial Irradiation

Cranial irradiation may result in long-term hypothalamic and pituitary dysfunction, especially in children and adolescents, as they are more susceptible to damage after whole-brain or head and neck therapeutic irradiation. The development of hormonal abnormalities correlates strongly with irradiation dosage and the time interval after completion of radiotherapy. Up to two-thirds of patients ultimately develop hormone

insufficiency after a median dose of 50 Gy (5000 rad) directed at the skull base. The development of hypopituitarism occurs over 5–15 years and usually reflects hypothalamic damage rather than primary destruction of pituitary cells. Although the pattern of hormone loss is variable, GH deficiency is most common, followed by gonadotropin and ACTH deficiency. When deficiency of one or more hormones is documented, the possibility of diminished reserve of other hormones is likely. Accordingly, anterior pituitary function should be continually evaluated over the long term in previously irradiated patients, and replacement therapy instituted when appropriate (see below).

Lymphocytic Hypophysitis

This occurs most often in postpartum women; it usually presents with hyperprolactinemia and MRI evidence of a prominent pituitary mass that often resembles an adenoma, with mildly elevated PRL levels. Pituitary failure caused by diffuse lymphocytic infiltration may be transient or permanent but requires immediate evaluation and treatment. Rarely, isolated pituitary hormone deficiencies have been described, suggesting a selective autoimmune process targeted to specific cell types. Most patients manifest symptoms of progressive mass effects with headache and visual disturbance. The erythrocyte sedimentation rate often is elevated. Because the MRI image may be indistinguishable from that of a pituitary adenoma, hypophysitis should be considered in a postpartum woman with a newly diagnosed pituitary mass before an unnecessary surgical intervention is undertaken. The inflammatory process often resolves after several months of glucocorticoid treatment, and pituitary function may be restored, depending on the extent of damage.

Immunotherapy and Hypophysitis

Pituitary cells express cytotoxic T lymphocyte antigen-4 (CTLA-4) and up to 20% of patients receiving cancer immunotherapy with CTLA-4 blockers (e.g., [ipilimumab](#)) may develop hypophysitis with associated thyroid adrenal and gonadal failure. Pituitary hormone replacement, with or without high-dose glucocorticoids, may be safely tolerated with continued immunotherapy.

Pituitary Apoplexy

Acute intrapituitary hemorrhagic vascular events can cause substantial damage to the pituitary and surrounding sellar structures. Pituitary apoplexy may occur spontaneously in a preexisting adenoma; postpartum (Sheehan's syndrome); or in association with diabetes, hypertension, sickle cell anemia, or acute shock. The hyperplastic enlargement of the pituitary, which occurs normally during pregnancy, increases the risk for hemorrhage and infarction. Apoplexy is an endocrine emergency that may result in severe hypoglycemia, hypotension and shock, central nervous system (CNS) hemorrhage, and death. Acute symptoms may include severe headache with signs of meningeal irritation, bilateral visual changes, ophthalmoplegia, and, in severe cases, cardiovascular collapse and loss of consciousness. Pituitary computed tomography (CT) or MRI may reveal signs of intratumoral or sellar hemorrhage, with pituitary stalk deviation and compression of pituitary tissue.

Patients with no evident visual loss or impaired consciousness can be observed and managed conservatively with high-dose glucocorticoids. Those with significant or progressive visual loss, cranial nerve palsy, or loss of consciousness require urgent surgical decompression. Visual recovery after sellar surgery is inversely correlated with the length of time after the acute event. Therefore, severe ophthalmoplegia or visual deficits are indications for early surgery. Hypopituitarism is common after apoplexy.

Empty Sella

A partial or apparently totally empty sella is often an incidental MRI finding, and may be associated with intracranial hypertension. These patients usually have normal pituitary function, implying that the surrounding rim of pituitary tissue is fully functional. Hypopituitarism, however, may develop insidiously. Pituitary masses also may undergo clinically silent infarction and involution with development of a partial or totally empty sella by cerebrospinal fluid (CSF) filling the dural herniation. Rarely, small but functional pituitary adenomas may arise within the rim of normal pituitary tissue, and they are not always visible on MRI.

PRESENTATION AND DIAGNOSIS

The clinical manifestations of hypopituitarism depend on which hormones are lost and the extent of the hormone deficiency. GH deficiency causes growth disorders in children and leads to abnormal body composition in adults (see below). Gonadotropin deficiency causes menstrual disorders and infertility in women and decreased sexual function, infertility, and loss of secondary sexual characteristics in men. TSH and ACTH deficiencies usually develop later in the course of pituitary failure. TSH deficiency causes growth retardation in children and features of hypothyroidism in children and adults. The secondary form of adrenal insufficiency caused by ACTH deficiency leads to hypocortisolism with relative preservation of mineralocorticoid production. PRL deficiency causes failure of lactation. When lesions involve the posterior pituitary, polyuria and polydipsia reflect loss of vasopressin secretion. In patients with long-standing pituitary damage, epidemiologic studies document an increased mortality rate, primarily from increased cardiovascular and cerebrovascular disease. Previous head or neck irradiation is also a determinant of increased mortality rates in patients with hypopituitarism, especially from cerebrovascular disease.

LABORATORY INVESTIGATION

Biochemical diagnosis of pituitary insufficiency is made by demonstrating low levels of respective pituitary trophic hormones in the setting of low levels of target hormones. For example, low free thyroxine in the setting of a low or inappropriately normal TSH level suggests secondary hypothyroidism. Similarly, a low testosterone level without elevation of gonadotropins suggests hypogonadotropic hypogonadism. Provocative tests may be required to assess pituitary reserve ([Table 372-2](#)). GH responses to insulin-induced hypoglycemia, arginine, L-dopa, growth hormone–releasing hormone (GHRH), or growth hormone–releasing peptides (GHRPs) can be used to assess GH reserve. Corticotropin-releasing hormone (CRH) administration induces ACTH release, and administration of synthetic ACTH (cosyntropin) evokes adrenal cortisol release as

an indirect indicator of pituitary ACTH reserve (**Chap. 379**). ACTH reserve is most reliably assessed by measuring ACTH and cortisol levels during insulin-induced hypoglycemia. However, this test should be performed cautiously in patients with suspected adrenal insufficiency because of enhanced susceptibility to hypoglycemia and hypotension. Administering insulin to induce hypoglycemia is contraindicated in patients with active coronary artery disease or known seizure disorders.

TABLE 372-2

Tests of Pituitary Sufficiency

HORMONE	TEST	BLOOD SAMPLES	INTERPRETATION
Growth hormone (GH)	Insulin tolerance test: Regular insulin (0.05–0.15 U/kg IV)	–30, 0, 30, 60, 120 min for glucose and GH	Glucose <40 mg/dL; GH should be >3 µg/L
	GHRH test: 1 µg/kg IV	0, 15, 30, 45, 60, 120 min for GH	Normal response is GH >3 µg/L
	L-Arginine test: 30 g IV over 30 min	0, 30, 60, 120 min for GH	Normal response is GH >3 µg/L
	L-Dopa test: 500 mg PO	0, 30, 60, 120 min for GH	Normal response is GH >3 µg/L
Prolactin	TRH test: 200–500 µg IV	0, 20, and 60 min for TSH and PRL	Normal prolactin is >2 µg/L and increase >200% of baseline
ACTH	Insulin tolerance test: regular insulin (0.05–0.15 U/kg IV)	–30, 0, 30, 60, 90 min for glucose and cortisol	Glucose <40 mg/dL Cortisol should increase by >7 µg/dL or to >20 µg/dL
	CRH test: 1 µg/kg ovine CRH IV at 8 A.M.	0, 15, 30, 60, 90, 120 min for ACTH and cortisol	Basal ACTH increases 2- to 4-fold and peaks at 20–100 pg/mL Cortisol levels >20–25 µg/dL
	Metyrapone test: Metyrapone (30 mg/kg) at midnight	Plasma 11-deoxycortisol and cortisol at 8 A.M.; ACTH can also be measured	Plasma cortisol should be <4 g/dL to assure an adequate response Normal response is 11-deoxycortisol >7.5 µg/dL or ACTH >75 pg/mL
	Standard ACTH stimulation test: ACTH 1-24 (cosyntropin), 0.25 mg IM or IV	0, 30, 60 min for cortisol and aldosterone	Normal response is cortisol >21 g/dL and aldosterone response of >4 ng/dL above baseline

HORMONE	TEST	BLOOD SAMPLES	INTERPRETATION
	Low-dose ACTH test: ACTH 1-24 (cosyntropin), 1 µg IV	0, 30, 60 min for cortisol	Cortisol should be >21 g/dL
	3-day ACTH stimulation test consists of 0.25 mg ACTH 1-24 given IV over 8 h each day		Cortisol >21 g/dL
TSH	Basal thyroid function tests: T ₄ , T ₃ , TSH	Basal measurements	Low free thyroid hormone levels in the setting of TSH levels that are not appropriately increased indicate pituitary insufficiency
	TRH test: 200–500 µg IV	0, 20, 60 min for TSH and PRL ^a	TSH should increase by >5 mU/L unless thyroid hormone levels are increased
LH, FSH	LH, FSH, testosterone, estrogen	Basal measurements	Basal LH and FSH should be increased in postmenopausal women
			Low testosterone levels in the setting of low LH and FSH indicate pituitary insufficiency
	GnRH test: GnRH (100 µg) IV	0, 30, 60 min for LH and FSH	In most adults, LH should increase by 10 IU/L and FSH by 2 IU/L
			Normal responses are variable
Multiple hormones	Combined anterior pituitary test: GHRH (1 g/kg), CRH (1 µg/kg), GnRH (100 g), TRH (200 µg) are given IV	–30, 0, 15, 30, 60, 90, 120 min for GH, ACTH, cortisol, LH, FSH, and TSH	Combined or individual releasing hormone responses must be elevated in the context of basal target gland hormone values and may not be uniformly diagnostic (see text)

^aEvoked PRL response indicates lactotrope integrity.

Abbreviations: T₃, triiodothyronine; T₄, thyroxine; TRH, thyrotropin-releasing hormone. For other abbreviations, see text.

TREATMENT

TREATMENT

Hypopituitarism

Hormone replacement therapy, including glucocorticoids, thyroid hormone, sex steroids, GH, and vasopressin, is usually safe and free of complications. Treatment regimens that mimic physiologic hormone production allow for maintenance of satisfactory clinical homeostasis. Effective dosage schedules are outlined in [Table 372-3](#). Patients in need of glucocorticoid replacement require careful dose adjustments during stressful events such as acute illness, dental procedures, trauma, and acute hospitalization.

TABLE 372-3

Hormone Replacement Therapy for Adult Hypopituitarism^a

TROPIC HORMONE DEFICIT	HORMONE REPLACEMENT
ACTH	Hydrocortisone (10–20 mg/d in divided doses)
	Cortisone acetate (15–25 mg/d in divided doses)
	Prednisone (5 mg A.M.)
TSH	L-Thyroxine (0.075–0.15 mg daily)
FSH/LH	Males
	Testosterone gel (5–10 g/d)
	Testosterone skin patch (5 mg/d) Testosterone enanthate (200 mg IM every 2 weeks)
	Females
	Conjugated estrogen (0.65–1.25 mg qd for 25 days)
	Progesterone (5–10 mg qd) on days 16–25
	Estradiol skin patch (0.025–0.1 mg every week), adding progesterone on days 16–25 if uterus intact
	For fertility: menopausal gonadotropins, human chorionic gonadotropins
GH	Adults: Somatotropin (0.1–1.25 mg SC qd)
	Children: Somatotropin (0.02–0.05 mg/kg per day)
Vasopressin	Intranasal desmopressin (5–20 g twice daily) Oral 300–600 µg qd

^aAll doses shown should be individualized for specific patients and should be reassessed during stress, surgery, or pregnancy. Male and female fertility requirements should be managed as discussed in [Chaps. 384 and 385](#).

Note: For abbreviations, see text.

DISORDERS OF GROWTH AND DEVELOPMENT

Skeletal Maturation and Somatic Growth

The growth plate is dependent on a variety of hormonal stimuli, including GH, insulin-like growth factor (IGF)-I, sex steroids, thyroid hormones, paracrine growth factors, and cytokines. The growth-promoting process also requires caloric energy, amino acids, vitamins, and trace metals and consumes about 10% of normal energy production. Malnutrition impairs chondrocyte activity, increases GH resistance, and reduces circulating IGF-I and IGF binding protein (IGBP)-3 levels.

Linear bone growth rates are very high in infancy and are pituitary-dependent. Mean growth velocity is ~6 cm/year in later childhood and usually is maintained within a given range on a standardized percentile chart. Peak growth rates occur during midpuberty when bone age is 12 (girls) or 13 (boys). Secondary sexual development is associated with elevated sex steroids that cause progressive epiphyseal growth plate closure. *Bone age* is delayed in patients with all forms of true GH deficiency or GH receptor defects that result in attenuated GH action.

Short stature may occur as a result of constitutive intrinsic growth defects or because of acquired extrinsic factors that impair growth. In general, delayed bone age in a child with short stature is suggestive of a hormonal or systemic disorder, whereas normal bone age in a short child is more likely to be caused by a genetic cartilage dysplasia or growth plate disorder ([Chap. 406](#)).

GH Deficiency in Children

GH DEFICIENCY

Isolated GH deficiency is characterized by short stature, micropenis, increased fat, high-pitched voice, and a propensity to hypoglycemia due to relatively unopposed insulin action. Familial modes of inheritance are seen in at least one-third of these individuals and may be autosomal dominant, recessive, or X-linked. About 10% of children with GH deficiency have mutations in the *GH-N* gene, including gene deletions and a wide range of point mutations. Mutations in transcription factors Pit-1 and Prop-1, which control somatotrope development, result in GH deficiency in combination with other pituitary hormone deficiencies, which may become manifest only in adulthood. The diagnosis of *idiopathic GH deficiency* (IGHD) should be made only after known molecular defects have been rigorously excluded.

GHRH RECEPTOR MUTATIONS

Recessive mutations of the GHRH receptor gene in subjects with severe proportionate dwarfism are associated with low basal GH levels that cannot be stimulated by exogenous GHRH, GHRP, or insulin-induced

hypoglycemia, as well as anterior pituitary hypoplasia. The syndrome exemplifies the importance of the GHRH receptor for somatotrope cell proliferation and hormonal responsiveness.

GH INSENSITIVITY

This is caused by defects of GH receptor structure or signaling. Homozygous or heterozygous mutations of the GH receptor are associated with partial or complete GH insensitivity and growth failure (*Laron syndrome*). The diagnosis is based on normal or high GH levels, with decreased circulating GH-binding protein (GHBP), and low IGF-I levels. Very rarely, defective IGF-I, IGF-I receptor, or IGF-I signaling defects are also encountered. *STAT5B* mutations result in both immunodeficiency as well as abrogated GH signaling, leading to short stature with normal or elevated GH levels and low IGF-I levels. Circulating GH receptor antibodies may rarely cause peripheral GH insensitivity.

NUTRITIONAL SHORT STATURE

Caloric deprivation and malnutrition, uncontrolled diabetes, and chronic renal failure represent secondary causes of abrogated GH receptor function. These conditions also stimulate production of proinflammatory cytokines, which act to exacerbate the block of GH-mediated signal transduction. Children with these conditions typically exhibit features of acquired short stature with normal or elevated GH and low IGF-I levels.

PSYCHOSOCIAL SHORT STATURE

Emotional and social deprivation lead to growth retardation accompanied by delayed speech, discordant hyperphagia, and an attenuated response to administered GH. A nurturing environment restores growth rates.

PRESENTATION AND DIAGNOSIS

Short stature is commonly encountered in clinical practice, and the decision to evaluate these children requires clinical judgment in association with auxologic data and family history. Short stature should be evaluated comprehensively if a patient's height is >3 standard deviations (SD) below the mean for age or if the growth rate has decelerated. Skeletal maturation is best evaluated by measuring a radiologic bone age, which is based mainly on the degree of wrist bone growth plate fusion. Final height can be predicted using standardized scales (Bayley-Pinneau or Tanner-Whitehouse) or estimated by adding 6.5 cm (boys) or subtracting 6.5 cm (girls) from the midparental height.

LABORATORY INVESTIGATION

Because GH secretion is pulsatile, GH deficiency is best assessed by examining the response to provocative stimuli, including exercise, insulin-induced hypoglycemia, and other pharmacologic tests that normally increase GH to >7 $\mu\text{g/L}$ in children. Random GH measurements do not distinguish normal children from those with true GH deficiency. Adequate adrenal and thyroid hormone replacement should be assured before testing. Age- and sex-matched IGF-I levels are not sufficiently sensitive or specific to make the diagnosis but can be useful to confirm GH deficiency. Pituitary MRI may reveal pituitary mass lesions or structural defects.

Molecular analyses for known mutations should be undertaken when the cause of short stature remains cryptic, or when additional clinical features suggest a genetic cause.

TREATMENT

TREATMENT

Disorders of Growth and Development

Replacement therapy with recombinant GH (0.02–0.05 mg/kg per day SC) restores growth velocity in GH-deficient children to ~10 cm/year. If pituitary insufficiency is documented, other associated hormone deficits should be corrected, especially adrenal steroids. GH treatment is also moderately effective for accelerating growth rates in children with Turner syndrome and chronic renal failure.

In patients with GH insensitivity and growth retardation due to mutations of the GH receptor, treatment with IGF-I bypasses the dysfunctional GH receptor.

ADULT GH DEFICIENCY (AGHD)

This disorder usually is caused by acquired hypothalamic or pituitary somatotrope damage. Acquired pituitary hormone deficiency follows a typical pattern in which loss of adequate GH reserve foreshadows subsequent hormone deficits. The sequential order of hormone loss is usually GH → FSH/LH → TSH → ACTH. Patients previously diagnosed with childhood-onset GH deficiency should be retested as adults to affirm the diagnosis.

PRESENTATION AND DIAGNOSIS

The clinical features of AGHD include changes in body composition, lipid metabolism, and quality of life and cardiovascular dysfunction ([Table 372-4](#)). Body composition changes are common and include reduced lean body mass, increased fat mass with selective deposition of intraabdominal visceral fat, and increased waist-to-hip ratio. Hyperlipidemia, left ventricular dysfunction, hypertension, and increased plasma fibrinogen levels also may be present. Bone mineral content is reduced, with resultant increased fracture rates. Patients may experience social isolation, depression, and difficulty maintaining gainful employment. Adult hypopituitarism is associated with a threefold increase in cardiovascular mortality rates in comparison to age- and sex-matched controls, and this may be due to GH deficiency, as patients in these studies were replaced with other deficient pituitary hormones.

TABLE 372-4

Features of Adult Growth Hormone Deficiency

Clinical
<p>Impaired quality of life</p> <ul style="list-style-type: none"> Decreased energy and drive Poor concentration Low self-esteem Social isolation <p>Body composition changes</p> <ul style="list-style-type: none"> Increased body fat mass Central fat deposition Increased waist-to-hip ratio Decreased lean body mass <p>Reduced exercise capacity</p> <ul style="list-style-type: none"> Reduced maximum O₂ uptake Impaired cardiac function Reduced muscle mass <p>Cardiovascular risk factors</p> <ul style="list-style-type: none"> Impaired cardiac structure and function Abnormal lipid profile Decreased fibrinolytic activity Atherosclerosis Omental obesity
Imaging
<p>Pituitary: mass or structural damage</p> <p>Bone: reduced bone mineral density</p> <p>Abdomen: excess omental adiposity</p>
Laboratory
<p>Evoked GH <3 ng/mL</p> <p>IGF-I and IGFBP3 low or normal</p> <p>Increased LDL cholesterol</p> <p>Concomitant gonadotropin, TSH, and/or ACTH reserve deficits may be present</p>

Abbreviation: LDL, low-density lipoprotein. For other abbreviations, see text.

LABORATORY INVESTIGATION

AGHD is rare, and in light of the nonspecific nature of associated clinical symptoms, patients appropriate for testing should be selected carefully on the basis of well-defined criteria. With few exceptions, testing should be restricted to patients with the following predisposing factors: (1) pituitary surgery, (2) pituitary or hypothalamic tumor or granulomas, (3) history of cranial irradiation, (4) radiologic evidence of a pituitary lesion, and (5) childhood requirement for GH replacement therapy. The transition of a GH-deficient adolescent to adulthood requires retesting to document subsequent AGHD. Up to 20% of patients previously treated for childhood-onset GH deficiency are found to be GH-sufficient on repeat testing as adults.

A significant proportion (~25%) of truly GH-deficient adults have low-normal IGF-I levels. Thus, as in the evaluation of GH deficiency in children, valid age- and sex-matched IGF-I measurements provide a useful index of therapeutic responses but are not sufficiently sensitive for diagnostic purposes. The most validated test to distinguish pituitary-sufficient patients from those with AGHD is insulin-induced (0.05–0.1 U/kg) hypoglycemia. After glucose reduction to ~40 mg/dL, most individuals experience neuroglycopenic symptoms (**Chap. 399**), and peak GH release occurs at 60 min and remains elevated for up to 2 h. About 90% of healthy adults exhibit GH responses >5 µg/L; AGHD is defined by a peak GH response to hypoglycemia of <3 µg/L. Although insulin-induced hypoglycemia is safe when performed under appropriate supervision, it is contraindicated in patients with diabetes, ischemic heart disease, cerebrovascular disease, or epilepsy and in elderly patients. Alternative stimulatory tests include intravenous arginine (30 g), GHRH (1 µg/kg), GHRP-6 (90 µg), and glucagon (1 mg). Combinations of these tests may evoke GH secretion in subjects who are not responsive to a single test.

TREATMENT

TREATMENT

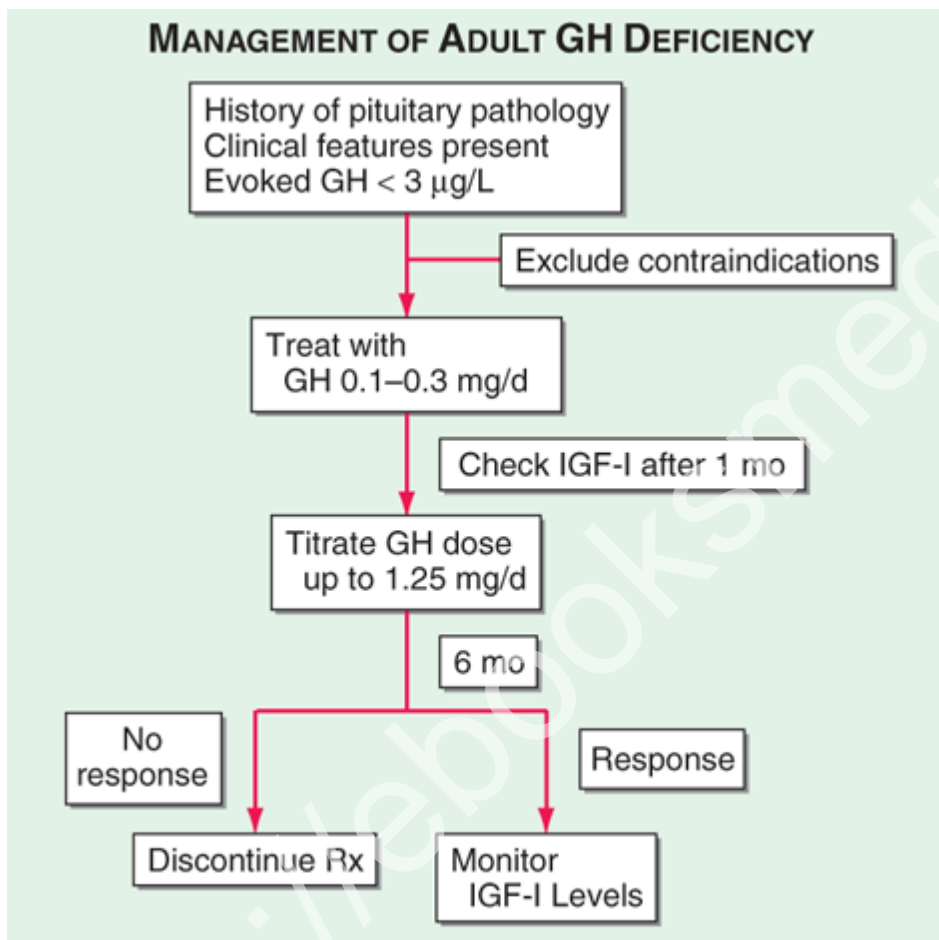
Adult GH Deficiency

Once the diagnosis of AGHD is unequivocally established, replacement of GH may be indicated. Contraindications to therapy include the presence of an active neoplasm, intracranial hypertension, and uncontrolled diabetes and retinopathy. The starting adult dose of 0.1–0.2 mg/d should be titrated (up to a maximum of 1.25 mg/d) to maintain IGF-I levels in the mid-normal range for age- and sex-matched controls (**Fig. 372-1**). Women require higher doses than men, and elderly patients require less GH. Long-term GH maintenance sustains normal IGF-I levels and is associated with persistent body composition changes (e.g., enhanced lean body mass and lower body fat). High-density lipoprotein cholesterol increases, but total cholesterol and insulin levels may not change significantly. Lumbar spine bone mineral density increases, but this response is gradual (>1 year). Many patients note significant improvement in quality of life when evaluated by standardized questionnaires. The effect of GH replacement on mortality rates in GH-deficient patients is currently the subject of long-term prospective investigation.

About 30% of patients exhibit reversible dose-related fluid retention, joint pain, and carpal tunnel syndrome, and up to 40% exhibit myalgias and paresthesia. Patients receiving insulin require careful monitoring for dosing adjustments, as GH is a potent counterregulatory hormone for insulin action. Patients with type 2 diabetes mellitus may initially develop further insulin resistance. However, glycemic control usually improves with the sustained loss of abdominal fat associated with long-term GH replacement. Headache, increased intracranial pressure, hypertension, and tinnitus occur rarely. Pituitary tumor regrowth and progression of skin lesions or other tumors have not been encountered in long-term surveillance programs with appropriate replacement doses.

FIGURE 372-1

Management of adult growth hormone (GH) deficiency. IGF, insulin-like growth factor; Rx, Treatment.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

ACTH DEFICIENCY

PRESENTATION AND DIAGNOSIS

Secondary adrenal insufficiency occurs as a result of pituitary ACTH deficiency. It is characterized by fatigue, weakness, anorexia, nausea, vomiting, and, occasionally, hypoglycemia. In contrast to primary adrenal

failure, hypocortisolism associated with pituitary failure usually is not accompanied by hyperpigmentation or mineralocorticoid deficiency.

ACTH deficiency is commonly due to glucocorticoid withdrawal after treatment-associated suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Isolated ACTH deficiency may occur after surgical resection of an ACTH-secreting pituitary adenoma that has suppressed the HPA axis; this phenomenon is in fact suggestive of a surgical cure. The mass effects of other pituitary adenomas or sellar lesions may lead to ACTH deficiency, usually in combination with other pituitary hormone deficiencies. Partial ACTH deficiency may be unmasked in the presence of an acute medical or surgical illness, when clinically significant hypocortisolism reflects diminished ACTH reserve. Rarely, *TPIT* or *POMC* mutations result in primary ACTH deficiency.

LABORATORY DIAGNOSIS

Inappropriately low ACTH levels in the setting of low cortisol levels are characteristic of diminished ACTH reserve. Low basal serum cortisol levels are associated with blunted cortisol responses to ACTH stimulation and impaired cortisol response to insulin-induced hypoglycemia, or testing with [metyrapone](#) or CRH. **For a description of provocative ACTH tests, see [Chap. 379](#).**

TREATMENT

TREATMENT

ACTH Deficiency

Glucocorticoid replacement therapy improves most features of ACTH deficiency. The total daily dose of [hydrocortisone](#) replacement preferably should generally not exceed 20 mg daily, divided into two or three doses. [Prednisone](#) (5 mg each morning) is longer acting and has fewer mineralocorticoid effects than [hydrocortisone](#). Some authorities advocate lower maintenance doses in an effort to avoid cushingoid side effects. Doses should be increased severalfold during periods of acute illness or stress. Patients should wear medical alert bracelets and/or carry identification cards with information about their glucocorticoid requirements.

GONADOTROPIN DEFICIENCY

Hypogonadism is the most common presenting feature of adult hypopituitarism even when other pituitary hormones are also deficient. It is often a harbinger of hypothalamic or pituitary lesions that impair GnRH production or delivery through the pituitary stalk. As noted below, hypogonadotropic hypogonadism is a common presenting feature of hyperprolactinemia.

A variety of inherited and acquired disorders are associated with *isolated hypogonadotropic hypogonadism* (IHH) ([Chap. 384](#)). Hypothalamic defects associated with GnRH deficiency include Kallmann syndrome and mutations in more than a dozen genes that regulate GnRH neuron migration, development, and function (see above). Mutations in *GPR54*, *DAX1*, kisspeptin, the GnRH receptor, and the LH β or FSH β subunit genes also

cause pituitary gonadotropin deficiency. Acquired forms of GnRH deficiency leading to hypogonadotropism are seen in association with anorexia nervosa, stress, starvation, and extreme exercise but also may be idiopathic. Hypogonadotropic hypogonadism in these disorders is reversed by removal of the stressful stimulus or by caloric replenishment.

PRESENTATION AND DIAGNOSIS

In premenopausal women, hypogonadotropic hypogonadism presents as diminished ovarian function leading to oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, and breast atrophy. In hypogonadal adult men, secondary testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, reduced beard and body hair growth, soft testes, and characteristic fine facial wrinkles. Osteoporosis occurs in both untreated hypogonadal women and men.

LABORATORY INVESTIGATION

Central hypogonadism is associated with low or inappropriately normal serum gonadotropin levels in the setting of low sex hormone concentrations (testosterone in men, [estradiol](#) in women). Because gonadotropin secretion is pulsatile, valid assessments may require repeated measurements or the use of pooled serum samples. Men have reduced sperm counts.

Intravenous GnRH (100 µg) stimulates gonadotropes to secrete LH (which peaks within 30 min) and FSH (which plateaus during the ensuing 60 min). Normal responses vary according to menstrual cycle stage, age, and sex of the patient. Generally, LH levels increase about threefold, whereas FSH responses are less pronounced. In the setting of gonadotropin deficiency, a normal gonadotropin response to GnRH indicates intact pituitary gonadotrope function and suggests a hypothalamic abnormality. An absent response, however, does not reliably distinguish pituitary from hypothalamic causes of hypogonadism. For this reason, GnRH testing usually adds little to the information gained from baseline evaluation of the hypothalamic-pituitary-gonadotrope axis except in cases of isolated GnRH deficiency (e.g., Kallmann syndrome).

MRI examination of the sellar region and assessment of other pituitary functions usually are indicated in patients with documented central hypogonadism.

TREATMENT

TREATMENT

Gonadotropin Deficiency

In males, testosterone replacement is necessary to achieve and maintain normal growth and development of the external genitalia, secondary sex characteristics, male sexual behavior, and androgenic anabolic effects, including maintenance of muscle function and bone mass. Testosterone may be administered by intramuscular injections every 1–4 weeks or by using skin patches or testosterone gels ([Chap. 384](#)).

Gonadotropin injections (hCG or human menopausal gonadotropin [hMG]) over 12–18 months are used to restore fertility. Pulsatile GnRH therapy (25–150 ng/kg every 2 h), administered by a subcutaneous infusion pump, is also effective for treatment of hypothalamic hypogonadism when fertility is desired.

In premenopausal women, cyclical replacement of estrogen and progesterone maintains secondary sexual characteristics and integrity of genitourinary tract mucosa and prevents premature osteoporosis ([Chap. 385](#)). Gonadotropin therapy is used for ovulation induction. Follicular growth and maturation are initiated using hMG or recombinant FSH; hCG or human luteinizing hormone (hLH) is subsequently injected to induce ovulation. As in men, pulsatile GnRH therapy can be used to treat hypothalamic causes of gonadotropin deficiency.

DIABETES INSIPIDUS

See [Chap. 374](#) for diagnosis and treatment of diabetes insipidus.

FURTHER READING

Alatzoglou KS, Dattani MT: Genetic causes and treatment of isolated growth hormone deficiency—An update. *Nat Rev Endocrinol* 6:562, 2010. [[PubMed: 20852587](#)]

Clemmons DR: The diagnosis and treatment of growth hormone deficiency in adults. *Curr Opin Endocrinol Diabetes Obes* 17:377, 2010. [[PubMed: 20588115](#)]

Dekkers OM et al: Multisystem morbidity and mortality in Cushing's syndrome: A cohort study. *J Clin Endocrinol Metab* 98:2277, 2013. [[PubMed: 23533241](#)]

Fleseriu M et al: Hormonal replacement in hypopituitarism in adults: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 101:3888, 2016. [[PubMed: 27736313](#)]

Honegger J et al: Diagnosis of primary hypophysitis in Germany. *J Clin Endocrinol Metab* 100:3841, 2015. [[PubMed: 26262437](#)]

Romijn JA: The chronic syndromes after previous treatment of pituitary tumours. *Nat Rev Endocrinol* 12:547, 2016. [[PubMed: 27259177](#)]

Sherlock M et al: Mortality in patients with pituitary disease. *Endocr Rev* 31:301, 2010. [[PubMed: 20086217](#)]

Tanriverdi F et al: Pituitary dysfunction after traumatic brain injury: A clinical and pathophysiological approach. *Endocr Rev* 36:305, 2015. [[PubMed: 25950715](#)]

3/14/2018

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 373: Pituitary Tumor Syndromes

Shlomo Melmed; J. Larry Jameson

HYPOTHALAMIC, PITUITARY, AND OTHER SELLAR MASSES

EVALUATION OF SELLAR MASSES

Local Mass Effects

Clinical manifestations of sellar lesions vary, depending on the anatomic location of the mass and the direction of its extension ([Table 373-1](#)). The dorsal sellar diaphragm presents the least resistance to soft tissue expansion from the sella; consequently, pituitary adenomas frequently extend in a suprasellar direction. Bony invasion may occur as well.

TABLE 373-1

Features of Sellar Mass Lesions^a

IMPACTED STRUCTURE	CLINICAL IMPACT
Pituitary	Hypogonadism Hypothyroidism Growth failure and adult hyposomatotropism Hypoadrenalism
Optic chiasm	Loss of red perception Bitemporal hemianopia Superior or bitemporal field defect Scotoma Blindness
Hypothalamus	Temperature dysregulation Appetite and thirst disorders Obesity Diabetes insipidus Sleep disorders Behavioral dysfunction Autonomic dysfunction
Cavernous sinus	Ophthalmoplegia with or without ptosis or diplopia Facial numbness
Frontal lobe	Personality disorder Anosmia
Brain	Headache Hydrocephalus Psychosis Dementia Laughing seizures

^aAs the intrasellar mass expands, it first compresses intrasellar pituitary tissue, then usually invades dorsally through the dura to lift the optic chiasm or laterally to the cavernous sinuses. Bony erosion is rare, as is direct brain compression. Microadenomas may present with headache.

Headaches are common features of small intrasellar tumors, even with no demonstrable suprasellar extension. Because of the confined nature of the pituitary, small changes in intrasellar pressure stretch the dural plate; however, headache severity correlates poorly with adenoma size or extension.

Suprasellar extension can lead to visual loss by several mechanisms, the most common being compression of the optic chiasm, but rarely, direct invasion of the optic nerves or obstruction of cerebrospinal fluid (CSF) flow leading to secondary visual disturbances can occur. Pituitary stalk compression by a hormonally active or inactive intrasellar mass may compress the portal vessels, disrupting pituitary access to hypothalamic hormones and dopamine; this results in early hyperprolactinemia and later concurrent loss of other pituitary hormones. This “stalk section” phenomenon may also be caused by trauma, whiplash injury with posterior clinoid stalk compression, or skull base fractures. Lateral mass invasion may impinge on the cavernous sinus and compress its neural contents, leading to cranial nerve III, IV, and VI palsies as well as effects on the ophthalmic and maxillary branches of the fifth cranial nerve (**Chap. 433**). Patients may present with diplopia, ptosis, ophthalmoplegia, and decreased facial sensation, depending on the extent of neural damage. Extension into the sphenoid sinus indicates that the pituitary mass has eroded through the sellar floor. Aggressive tumors rarely invade the palate roof and cause nasopharyngeal obstruction, infection, and CSF leakage. Temporal and frontal lobe involvement may rarely lead to uncinate seizures, personality disorders, and anosmia. Direct hypothalamic encroachment by an invasive pituitary mass may cause important metabolic sequelae, including precocious puberty or hypogonadism, diabetes insipidus, sleep disturbances, dysthermia, and appetite disorders.

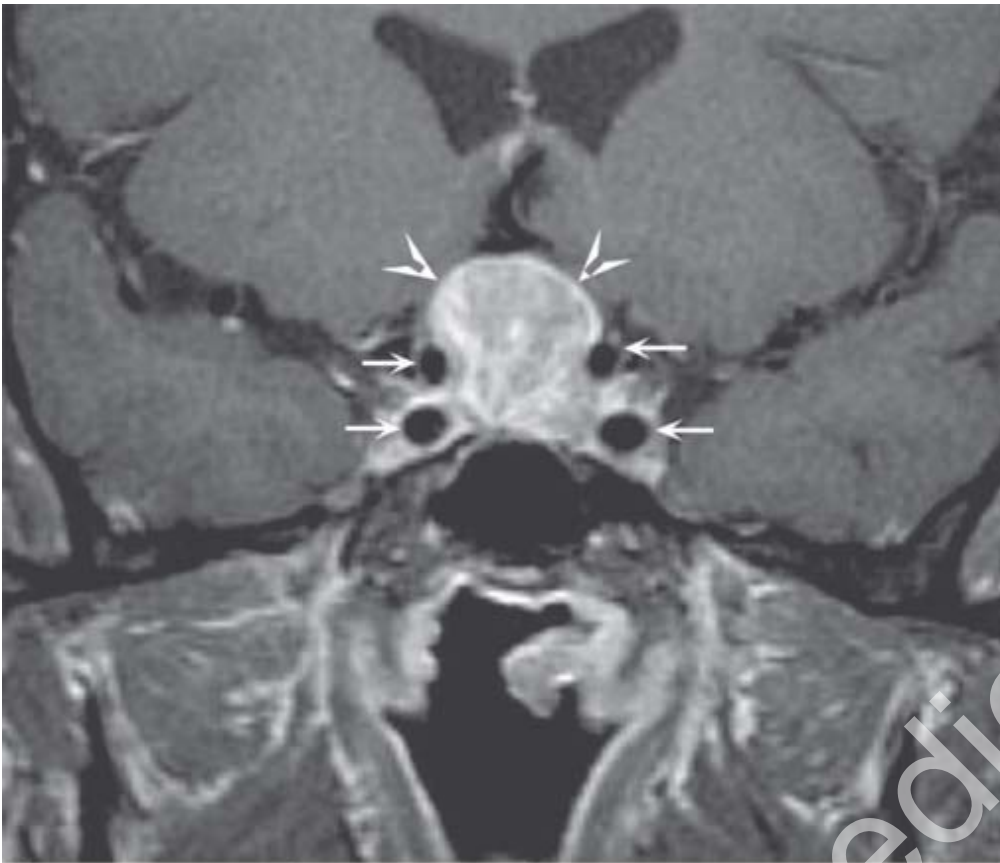
Magnetic Resonance Imaging

Sagittal and coronal T1-weighted magnetic resonance imaging (MRI) before and after administration of gadolinium allows precise visualization of the pituitary gland with clear delineation of the hypothalamus, pituitary stalk, pituitary tissue and surrounding suprasellar cisterns, cavernous sinuses, sphenoid sinus, and optic chiasm. Pituitary gland height ranges from 6 mm in children to 8 mm in adults; during pregnancy and puberty, the height may reach 10–12 mm. The upper aspect of the adult pituitary is flat or slightly concave, but in adolescent and pregnant individuals, this surface may be convex, reflecting physiologic pituitary enlargement. The stalk should be midline and vertical. Computed tomography (CT) scan is reserved to define the extent of bony erosion or the presence of calcification.

Anterior pituitary gland soft tissue consistency is slightly heterogeneous on MRI, and signal intensity resembles that of brain matter on T1-weighted imaging (**Fig. 373-1**). Adenoma density is usually lower than that of surrounding normal tissue on T1-weighted imaging, and the signal intensity increases with T2-weighted images. The high phospholipid content of the posterior pituitary results in a “pituitary bright spot.”

FIGURE 373-1

Pituitary adenoma. Coronal T1-weighted postcontrast magnetic resonance image shows a homogeneously enhancing mass (*arrowheads*) in the sella turcica and suprasellar region compatible with a pituitary adenoma; the *small arrows* outline the carotid arteries.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Sellar masses are encountered commonly as incidental findings on MRI, and most of them are pituitary adenomas (incidentalomas). In the absence of hormone hypersecretion, these small intrasellar lesions can be monitored safely with MRI, which is performed annually and then less often if there is no evidence of further growth. Resection should be considered for incidentally discovered larger macroadenomas, because about one-third become invasive or cause local pressure effects. If hormone hypersecretion is evident, specific therapies are indicated as described below. When larger masses (>1 cm) are encountered, they should also be distinguished from nonadenomatous lesions. Meningiomas often are associated with bony hyperostosis; craniopharyngiomas may be calcified and are usually hypodense, whereas gliomas are hyperdense on T2-weighted images.

Ophthalmologic Evaluation

Because optic tracts may be contiguous to an expanding pituitary mass, reproducible visual field assessment using perimetry techniques should be performed on all patients with sellar mass lesions that impinge the optic chiasm (**Chap. 28**). Bitemporal hemianopia, often more pronounced superiorly, is observed classically. It occurs because nasal ganglion cell fibers, which cross in the optic chiasm, are especially vulnerable to compression of the ventral optic chiasm. Occasionally, homonymous hemianopia occurs from postchiasmal compression or monocular temporal field loss from prechiasmal compression. Invasion of the cavernous sinus can produce diplopia from ocular motor nerve palsy. Early diagnosis reduces the risk of optic atrophy, vision loss, or eye misalignment.

Laboratory Investigation

The presenting clinical features of functional pituitary adenomas (e.g., acromegaly, prolactinomas, or Cushing syndrome) should guide the laboratory studies ([Table 373-2](#)). However, for a sellar mass with no obvious clinical features of hormone excess, laboratory studies are geared toward determining the nature of the tumor and assessing the possible presence of hypopituitarism. When a pituitary adenoma is suspected based on MRI, initial hormonal evaluation usually includes (1) basal prolactin (PRL); (2) insulin-like growth factor (IGF)-I; (3) 24-h urinary free cortisol (UFC) and/or overnight oral [dexamethasone](#) (1 mg) suppression test; (4) α subunit, follicle-stimulating hormone (FSH), and luteinizing hormone (LH); and (5) thyroid function tests. Additional hormonal evaluation may be indicated based on the results of these tests. Pending more detailed assessment of hypopituitarism, a menstrual history, measurement of testosterone and 8 A.M. cortisol levels, and thyroid function tests usually identify patients with pituitary hormone deficiencies that require hormone replacement before further testing or surgery.

TABLE 373-2

Screening Tests for Functional Pituitary Adenomas

	TEST	COMMENTS
Acromegaly	Serum IGF-I Oral glucose tolerance test with GH obtained at 0, 30, and 60 min	Interpret IGF-I relative to age- and sex-matched controls Normal subjects should suppress growth hormone to <1 g/L
Prolactinoma	Serum PRL	Exclude medications MRI of the sella should be ordered if PRL is elevated
Cushing's disease	24-h urinary free cortisol Dexamethasone (1 mg) at 11 P.M. and fasting plasma cortisol measured at 8 A.M. ACTH assay	Ensure urine collection is total and accurate Normal subjects suppress to <5 g/dL Distinguishes adrenal adenoma (ACTH suppressed) from ectopic ACTH or Cushing's disease (ACTH normal or elevated)

Abbreviations: ACTH, adrenocorticotropin hormone; GH, growth hormone; IGF-I, insulin-like growth factor I; MRI, magnetic resonance imaging; PRL, prolactin.

Histologic Evaluation

Immunohistochemical staining of pituitary tumor specimens obtained at transsphenoidal surgery confirms clinical and laboratory studies and provides a histologic diagnosis when hormone studies are equivocal and in cases of clinically nonfunctioning tumors.

TREATMENT

TREATMENT

Hypothalamic, Pituitary, and Other Sellar Masses

OVERVIEW

Successful management of sellar masses requires accurate diagnosis as well as selection of optimal therapeutic modalities. Most pituitary tumors are benign and slow-growing. Clinical features result from local mass effects and hormonal hyper- or hyposecretion syndromes caused directly by the adenoma or occurring as a consequence of treatment. Thus, lifelong management and follow-up are necessary for these patients.

MRI with gadolinium enhancement for pituitary visualization, new advances in transsphenoidal surgery and in stereotactic radiotherapy (including gamma-knife radiotherapy), and novel therapeutic agents have improved pituitary tumor management. The goals of pituitary tumor treatment include normalization of excess pituitary secretion, amelioration of symptoms and signs of hormonal hypersecretion syndromes, and shrinkage or ablation of large tumor masses with relief of adjacent structure compression. Residual anterior pituitary function should be preserved during treatment and sometimes can be restored by removing the tumor mass. Ideally, adenoma recurrence should be prevented.

TRANSSPHEOIDAL SURGERY

Transsphenoidal rather than transfrontal resection is the desired surgical approach for pituitary tumors, except for the rare invasive suprasellar mass surrounding the frontal or middle fossa or the optic nerves or invading posteriorly behind the clivus. Intraoperative microscopy facilitates visual distinction between adenomatous and normal pituitary tissue as well as microdissection of small tumors that may not be visible by MRI (**Fig. 373-2**). Transsphenoidal surgery also avoids the cranial invasion and manipulation of brain tissue required by subfrontal surgical approaches. Endoscopic techniques with three-dimensional intraoperative localization have improved visualization and access to tumor tissue. Individual surgical experience is a major determinant of outcome efficacy with these techniques.

In addition to correction of hormonal hypersecretion, pituitary surgery is indicated for mass lesions that impinge on surrounding structures. Surgical decompression and resection are required for an expanding pituitary mass, which may be asymptomatic or accompanied by persistent headache, progressive visual field defects, cranial nerve palsies, hydrocephalus, and, occasionally, intrapituitary hemorrhage and apoplexy. Transsphenoidal surgery sometimes is used for pituitary tissue biopsy to establish a histologic diagnosis. Whenever possible, the pituitary mass lesion should be selectively excised; normal pituitary tissue should be manipulated or resected only when critical for effective mass dissection. Nonselective hemihypophysectomy or total hypophysectomy may be indicated if no hypersecreting mass lesion is clearly discernible, multifocal lesions are present, or the remaining nontumorous pituitary tissue is obviously necrotic. This strategy, however, increases the likelihood of postoperative hypopituitarism and the need for lifelong hormone replacement.

Preoperative mass effects, including visual field defects and compromised pituitary function, may be reversed by surgery, particularly when the deficits are not long-standing. For large and invasive tumors, it is necessary to determine the optimal balance between maximal tumor resection and preservation of anterior pituitary function, especially for preserving growth and reproductive function in younger patients. Similarly, tumor invasion outside the sella is rarely amenable to surgical cure; the surgeon must judge the risk-versus-benefit ratio of extensive tumor resection.

Side Effects

Tumor size, the degree of invasiveness, and experience of the surgeon largely determine the incidence of surgical complications. Operative mortality rate is ~1%. Transient diabetes insipidus and hypopituitarism occur in up to 20% of patients. Permanent diabetes insipidus, cranial nerve damage, nasal septal perforation, or visual disturbances may be encountered in up to 10% of patients. CSF leaks occur in 4% of patients. Less common complications include carotid artery injury, loss of vision, hypothalamic damage, and meningitis. Permanent side effects are rare after surgery for microadenomas.

RADIATION

Radiation is used either as a primary therapy for pituitary or parasellar masses or, more commonly, as an adjunct to surgery or medical therapy. Focused megavoltage irradiation is achieved by precise MRI localization, using a high-voltage linear accelerator and accurate isocentric rotational arcing. A major determinant of accurate irradiation is reproduction of the patient's head position during multiple visits and maintenance of absolute head immobility. A total of <50 Gy (5000 rad) is given as 180-cGy (180-rad) fractions divided over ~6 weeks. Stereotactic radiosurgery delivers a large single high-energy dose from a cobalt-60 source (gamma knife), linear accelerator, or cyclotron. Long-term effects of gamma-knife surgery are unclear but appear to be similar to those encountered with conventional radiation. Proton beam therapy is available in some centers and provides concentrated radiation doses within a localized region.

The role of radiation therapy in pituitary tumor management depends on multiple factors, including the nature of the tumor, the age of the patient, and the availability of surgical and radiation expertise. Because of its relatively slow onset of action, radiation therapy is usually reserved for postsurgical management. As an adjuvant to surgery, radiation is used to treat residual tumor and in an attempt to prevent regrowth. Irradiation offers the only means for potentially ablating significant postoperative residual nonfunctioning tumor tissue. In contrast, PRL-, growth hormone (GH)-, and adrenocorticotropin hormone (ACTH)-secreting residual tumor tissues are amenable to medical therapy.

Side Effects

In the short term, radiation may cause transient nausea and weakness. Alopecia and loss of taste and smell may be more long-lasting. Failure of pituitary hormone synthesis is common in patients who have undergone head and neck or pituitary-directed irradiation. More than 50% of patients develop loss of GH, ACTH, thyroid-stimulating hormone (TSH), and/or gonadotropin secretion within 10 years, usually due to hypothalamic damage. Lifelong follow-up with testing of anterior pituitary hormone reserve is therefore required after radiation treatment. Optic nerve damage with impaired vision due to optic neuritis is reported in ~2% of patients who undergo pituitary irradiation. Cranial nerve damage is uncommon now that radiation doses are

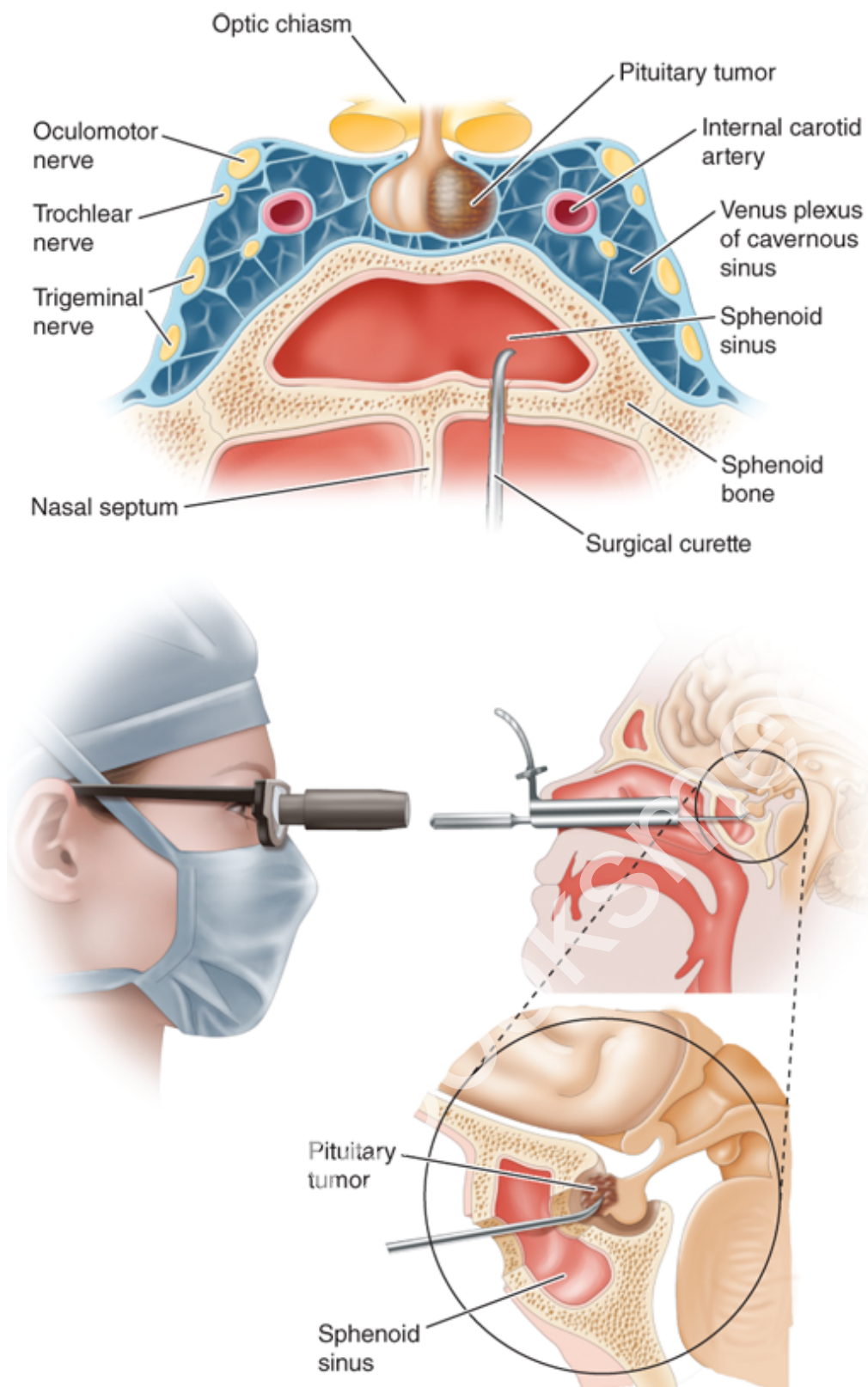
<2 Gy (200 rad) at any one treatment session and the maximum dose is <50 Gy (5000 rad). The use of stereotactic radiotherapy may reduce damage to adjacent structures. Radiotherapy for pituitary tumors has been associated with adverse mortality rates, mainly from cerebrovascular disease. The cumulative risk of developing a secondary tumor after conventional radiation is 1.3% after 10 years and 1.9% after 20 years.

MEDICAL

Medical therapy for pituitary tumors is highly specific and depends on tumor type. For prolactinomas, dopamine agonists are the treatment of choice. For acromegaly, somatostatin analogues and a GH receptor antagonist are indicated. For TSH-secreting tumors, somatostatin analogues and occasionally dopamine agonists are indicated. ACTH-secreting tumors may respond to somatostatin analogues, and adrenal-directed therapy may also be of benefit. Nonfunctioning tumors are generally not responsive to medications and require surgery and/or irradiation.

FIGURE 373-2

Transsphenoidal resection of pituitary mass via the endonasal approach. (*Adapted from R Fahlbusch: Endocrinol Metab Clin 21:669, 1992.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

SELLAR MASSES

Sellar masses other than pituitary adenomas may arise from brain, hypothalamic, or pituitary tissues. Each exhibit features related to the lesion location but also unique to the specific etiology.

Hypothalamic Lesions

Lesions involving the anterior and preoptic hypothalamic regions cause paradoxical vasoconstriction, tachycardia, and hyperthermia. Acute hyperthermia usually is due to a hemorrhagic insult, but poikilothermia may also occur. Central disorders of thermoregulation result from posterior hypothalamic damage. The *periodic hypothermia syndrome* is characterized by episodic attacks of rectal temperatures $<30^{\circ}\text{C}$ (86°F), sweating, vasodilation, vomiting, and bradycardia (**Chap. 454**). Damage to the ventromedial hypothalamic nuclei by craniopharyngiomas, hypothalamic trauma, or inflammatory disorders may be associated with *hyperphagia* and *obesity*. This region appears to contain an energy-satiety center where melanocortin receptors are influenced by leptin, insulin, pro-opiomelanocortin (POMC) products, and gastrointestinal peptides (**Chap. 394**). Polydipsia and hypodipsia are associated with damage to central osmoreceptors located in preoptic nuclei (**Chap. 374**). Slow-growing hypothalamic lesions can cause increased somnolence and disturbed sleep cycles as well as obesity, hypothermia, and emotional outbursts. Lesions of the central hypothalamus may stimulate sympathetic neurons, leading to elevated serum catecholamine and cortisol levels. These patients are predisposed to cardiac arrhythmias, hypertension, and gastric erosions.

Craniopharyngiomas are benign, suprasellar cystic masses that present with headaches, visual field deficits, and variable degrees of hypopituitarism. They are derived from Rathke's pouch and arise near the pituitary stalk, commonly extending into the suprasellar cistern. Craniopharyngiomas are often large, cystic, and locally invasive. Many are partially calcified, exhibiting a characteristic appearance on skull x-ray and CT images. More than half of all patients present before age 20, usually with signs of increased intracranial pressure, including headache, vomiting, papilledema, and hydrocephalus. Associated symptoms include visual field abnormalities, personality changes and cognitive deterioration, cranial nerve damage, sleep difficulties, and weight gain. Hypopituitarism can be documented in ~90%, and diabetes insipidus occurs in ~10% of patients. About half of affected children present with growth retardation. MRI is generally superior to CT for evaluating cystic structure and tissue components of craniopharyngiomas. CT is useful to define calcifications and evaluate invasion into surrounding bony structures and sinuses.

Treatment usually involves transcranial or transsphenoidal surgical resection followed by postoperative radiation of residual tumor. Surgery alone is curative in less than half of patients because of recurrences due to adherence to vital structures or because of small tumor deposits in the hypothalamus or brain parenchyma. The goal of surgery is to remove as much tumor as possible without risking complications associated with efforts to remove firmly adherent or inaccessible tissue. In the absence of radiotherapy, ~75% of craniopharyngiomas recur, and 10-year survival is $<50\%$. In patients with incomplete resection, radiotherapy improves 10-year survival to 70–90% but is associated with increased risk of secondary malignancies. Most patients require lifelong pituitary hormone replacement.

Developmental failure of Rathke's pouch obliteration may lead to *Rathke's cysts*, which are small ($<5\text{ mm}$) cysts entrapped by squamous epithelium and are found in ~20% of individuals at autopsy. Although Rathke's

cleft cysts do not usually grow and are often diagnosed incidentally, about a third present in adulthood with compressive symptoms, diabetes insipidus, and hyperprolactinemia due to stalk compression. Rarely, hydrocephalus develops. The diagnosis is suggested preoperatively by visualizing the cyst wall on MRI, which distinguishes these lesions from craniopharyngiomas. Cyst contents range from CSF-like fluid to mucoid material. *Arachnoid cysts* are rare and generate an MRI image that is isointense with CSF.

Sella chordomas usually present with bony clival erosion, local invasiveness, and, on occasion, calcification. Normal pituitary tissue may be visible on MRI, distinguishing chordomas from aggressive pituitary adenomas. Mucinous material may be obtained by fine-needle aspiration.

Meningiomas arising in the sellar region may be difficult to distinguish from nonfunctioning pituitary adenomas. Meningiomas typically enhance on MRI and may show evidence of calcification or bony erosion. Meningiomas may cause compressive symptoms.

Histiocytosis X includes a variety of syndromes associated with foci of eosinophilic granulomas. Diabetes insipidus, exophthalmos, and punched-out lytic bone lesions (*Hand-Schüller-Christian disease*) are associated with granulomatous lesions visible on MRI, as well as a characteristic axillary skin rash. Rarely, the pituitary stalk may be involved.

Pituitary metastases occur in ~3% of cancer patients. Bloodborne metastatic deposits are found almost exclusively in the posterior pituitary. Accordingly, diabetes insipidus can be a presenting feature of lung, gastrointestinal, breast, and other pituitary metastases. About half of pituitary metastases originate from breast cancer; ~25% of patients with metastatic breast cancer have such deposits. Rarely, pituitary stalk involvement results in anterior pituitary insufficiency. The MRI diagnosis of a metastatic lesion may be difficult to distinguish from an aggressive pituitary adenoma; the diagnosis may require histologic examination of excised tumor tissue. Primary or metastatic lymphoma, leukemias, and plasmacytomas also occur within the sella.

Hypothalamic hamartomas and *gangliocytomas* may arise from astrocytes, oligodendrocytes, and neurons with varying degrees of differentiation. These tumors may overexpress hypothalamic neuropeptides, including gonadotropin-releasing hormone (GnRH), growth hormone-releasing hormone (GHRH), and corticotropin-releasing hormone (CRH). With GnRH-producing tumors, children present with precocious puberty, psychomotor delay, and laughing-associated seizures. Medical treatment of GnRH-producing hamartomas with long-acting GnRH analogues effectively suppresses gonadotropin secretion and controls premature pubertal development. Rarely, hamartomas also are associated with craniofacial abnormalities; imperforate anus; cardiac, renal, and lung disorders; and pituitary failure as features of *Pallister-Hall syndrome*, which is caused by mutations in the carboxy terminus of the *GLI3* gene. Hypothalamic hamartomas are often contiguous with the pituitary, and preoperative MRI diagnosis may not be possible. Histologic evidence of hypothalamic neurons in tissue resected at transsphenoidal surgery may be the first indication of a primary hypothalamic lesion.

Hypothalamic gliomas and *optic gliomas* occur mainly in childhood and usually present with visual loss. Adults have more aggressive tumors; about a third are associated with neurofibromatosis.

Brain germ cell tumors may arise within the sellar region. They include *dysgerminomas*, which frequently are associated with diabetes insipidus and visual loss. They rarely metastasize. *Germinomas*, *embryonal carcinomas*, *teratomas*, and *choriocarcinomas* may arise in the parasellar region and produce hCG. These germ cell tumors present with precocious puberty, diabetes insipidus, visual field defects, and thirst disorders. Many patients are GH-deficient with short stature.

PITUITARY ADENOMAS AND HYPERSECRETION SYNDROMES

Pituitary adenomas are the most common cause of pituitary hormone hypersecretion and hyposecretion syndromes in adults. They account for ~15% of all intracranial neoplasms and have been identified with a population prevalence of ~80/100,000. At autopsy, up to one-quarter of all pituitary glands harbor an unsuspected microadenoma (<10 mm diameter). Similarly, pituitary imaging detects small clinically inapparent pituitary lesions in at least 10% of individuals.

Pathogenesis

Pituitary adenomas are benign neoplasms that arise from one of the five anterior pituitary cell types. The clinical and biochemical phenotypes of pituitary adenomas depend on the cell type from which they are derived. Thus, tumors arising from lactotrope (PRL), somatotrope (GH), corticotrope (ACTH), thyrotrope (TSH), or gonadotrope (LH, FSH) cells hypersecrete their respective hormones ([Table 373-3](#)). Plurihormonal tumors express various combinations of GH, PRL, TSH, ACTH, or the glycoprotein hormone α or β subunits. They may be diagnosed by careful immunocytochemistry or may manifest as clinical syndromes that combine features of these hormonal hypersecretory syndromes. Morphologically, these tumors may arise from a single polysecreting cell type or include cells with mixed function within the same tumor.

TABLE 373-3

Classification of Pituitary Adenomas^a

ADENOMA CELL ORIGIN	HORMONE PRODUCT	CLINICAL SYNDROME
Lactotrope	PRL	Hypogonadism, galactorrhea
Gonadotrope	FSH, LH, subunits	Silent or hypogonadism
Somatotrope	GH	Acromegaly/gigantism
Corticotrope	ACTH/none	Cushing's disease or silent
Mixed growth hormone and prolactin cell	GH, PRL	Acromegaly, hypogonadism, galactorrhea
Other plurihormonal cell	Any	Mixed
Acidophil stem cell	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Mammosomatotrope	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Thyrotrope	TSH	Thyrotoxicosis
Null cell	None	Pituitary failure/none
Oncocytoma	None	Pituitary failure/none

^aHormone-secreting tumors are listed in decreasing order of frequency. All tumors may cause local pressure effects, including visual disturbances, cranial nerve palsy, and headache.

Note: For abbreviations, see text.

Source: Adapted from S Melmed: Nat Rev Endocrinol 7:257, 2011.

Hormonally active tumors are characterized by autonomous hormone secretion with diminished feedback responsiveness to physiologic inhibitory pathways. Hormone production does not always correlate with tumor size. Small hormone-secreting adenomas may cause significant clinical perturbations, whereas larger adenomas that produce less hormone may be clinically silent and remain undiagnosed (if no central compressive effects occur). About one-third of all adenomas are clinically nonfunctioning and produce no distinct clinical hypersecretory syndrome. Most of them arise from gonadotrope cells and may secrete small

amounts of α - and β -glycoprotein hormone subunits or, very rarely, intact circulating gonadotropins. True pituitary carcinomas with documented extracranial metastases are exceedingly rare.

Almost all pituitary adenomas are monoclonal in origin, implying the acquisition of one or more somatic mutations that confer a selective growth advantage. Consistent with their clonal origin, complete surgical resection of small pituitary adenomas usually cures hormone hypersecretion. Nevertheless, hypothalamic hormones such as GHRH and CRH also enhance mitotic activity of their respective pituitary target cells in addition to their role in pituitary hormone regulation. Thus, patients who harbor rare abdominal or chest tumors that elaborate ectopic GHRH or CRH may present with somatotrope or corticotrope hyperplasia with GH or ACTH hypersecretion.

Several etiologic genetic events have been implicated in the development of pituitary tumors. The pathogenesis of sporadic forms of acromegaly has been particularly informative as a model of tumorigenesis. GHRH, after binding to its G protein–coupled somatotrope receptor, uses cyclic adenosine monophosphate (AMP) as a second messenger to stimulate GH secretion and somatotrope proliferation. A subset (~35%) of GH-secreting pituitary tumors contains sporadic mutations in *Gsa* (Arg 201 → Cys or His; Gln 227 → Arg). These mutations attenuate intrinsic GTPase activity, resulting in constitutive elevation of cyclic AMP, Pit-1 induction, and activation of cyclic AMP response element binding protein (CREB), thereby promoting somatotrope cell proliferation and GH secretion.

Characteristic loss of heterozygosity (LOH) in various chromosomes has been documented in large or invasive macroadenomas, suggesting the presence of putative tumor suppressor genes at these loci in up to 20% of sporadic pituitary tumors, including GH-, PRL-, and ACTH-producing adenomas and some nonfunctioning tumors. Lineage-specific cell cycle disruptions with elevated levels of CDK inhibitors are present in most pituitary adenomas.

Compelling evidence also favors growth factor promotion of pituitary tumor proliferation. Basic fibroblast growth factor (bFGF) is abundant in the pituitary and stimulates pituitary cell mitogenesis, whereas epithelial growth factor receptor (EGFR) signaling induces both hormone synthesis and cell proliferation. Mutations of *USP8* may result in overexpressed EGFR in a subset of ACTH-secreting tumors. Other factors involved in initiation and promotion of pituitary tumors include loss of negative-feedback inhibition (as seen with primary hypothyroidism or hypogonadism) and estrogen-mediated or paracrine angiogenesis. Growth characteristics and neoplastic behavior also may be influenced by several activated oncogenes, including *RAS* and pituitary tumor transforming gene (*PTTG*), or inactivation of growth suppressor genes, including *MEG3*.

Genetic Syndromes Associated with Pituitary Tumors

Several familial syndromes are associated with pituitary tumors, and the genetic mechanisms for some of them have been unraveled ([Table 373-4](#)).

TABLE 373-4

Familial Pituitary Tumor Syndromes

	GENE MUTATED	CLINICAL FEATURES
Multiple endocrine neoplasia 1 (MEN 1)	<i>MEN1</i> (11q13)	Hyperparathyroidism Pancreatic neuroendocrine tumors Foregut carcinoids Adrenal adenomas Skin lesions Pituitary adenomas (40%)
Multiple endocrine neoplasia 4 (MEN 4)	<i>CDKN1B</i> (12p13)	Hyperparathyroidism Pituitary adenomas Other tumors
Carney complex	<i>PRKAR1A</i> (17q23-24)	Pituitary hyperplasia and adenomas (10%) Atrial myxomas Schwannomas Adrenal hyperplasia Lentigines
Familial pituitary adenomas	<i>AIP</i> (11q.13.2)	Acromegaly/gigantism (~15% of afflicted families)

Multiple endocrine neoplasia (MEN) 1 is an autosomal dominant syndrome characterized primarily by a genetic predisposition to parathyroid, pancreatic islet, and pituitary adenomas (**Chap. 381**). MEN1 is caused by inactivating germline mutations in *MENIN*, a constitutively expressed tumor-suppressor gene located on chromosome 11q13. Loss of heterozygosity or a somatic mutation of the remaining normal *MENIN* allele leads to tumorigenesis. About half of affected patients develop prolactinomas; acromegaly and Cushing syndrome are less commonly encountered.

Carney complex is characterized by spotty skin pigmentation, myxomas, and endocrine tumors, including testicular, adrenal, and pituitary adenomas. Acromegaly occurs in ~20% of these patients. A subset of patients have mutations in the R1 α regulatory subunit of protein kinase A (*PRKAR1A*).

McCune-Albright syndrome consists of polyostotic fibrous dysplasia, pigmented skin patches, and a variety of endocrine disorders, including acromegaly, adrenal adenomas, and autonomous ovarian function (**Chap. 405**). Hormonal hypersecretion results from constitutive cyclic AMP production caused by inactivation of the

GTPase activity of Gs α . The Gs α mutations occur postzygotically, leading to a mosaic pattern of mutant expression.

Familial acromegaly is a rare disorder in which family members may manifest either acromegaly or gigantism. A subset of families with a predisposition for familial pituitary tumors, especially acromegaly, have been found to harbor germline mutations in the *AIP* gene, which encodes the aryl hydrocarbon receptor interacting protein.

HYPERPROLACTINEMIA

Etiology

Hyperprolactinemia is the most common pituitary hormone hypersecretion syndrome in both men and women. PRL-secreting pituitary adenomas (prolactinomas) are the most common cause of PRL levels >200 $\mu\text{g/L}$ (see below). Less pronounced PRL elevation can also be seen with microprolactinomas but is more commonly caused by drugs, pituitary stalk compression, hypothyroidism, or renal failure ([Table 373-5](#)).

TABLE 373-5

Etiology of Hyperprolactinemia**I. Physiologic hypersecretion**

Pregnancy
Lactation
Chest wall stimulation
Sleep
Stress

II. Hypothalamic–pituitary stalk damage

Tumors
 Craniopharyngioma
 Suprasellar pituitary mass
 Meningioma
 Dysgerminoma
 Metastases
Empty sella
Lymphocytic hypophysitis
Adenoma with stalk
Compression
Granulomas
Rathke cyst
Irradiation
Trauma
 Pituitary stalk section
 Suprasellar surgery

III. Pituitary hypersecretion

Prolactinoma
Acromegaly

IV. Systemic disorders

Chronic renal failure
Hypothyroidism
Cirrhosis
Pseudocyesis
Epileptic seizures

V. Drug-induced hypersecretion

Dopamine receptor blockers
 Atypical antipsychotics: risperidone
 Phenothiazines: chlorpromazine, perphenazine

Butyrophenones: haloperidol

Thioxanthenes

[Metoclopramide](#)

Dopamine synthesis inhibitors

α -Methyldopa

Catecholamine depletors

Reserpine

Opiates

H₂ antagonists

Cimetidine, [ranitidine](#)

Imipramines

Amitriptyline, amoxapine

Serotonin reuptake inhibitors

Fluoxetine

Calcium channel blockers

[Verapamil](#)

[Estrogens](#)

Thyrotropin-releasing hormone

Note: Hyperprolactinemia >200 µg/L almost invariably is indicative of a prolactin-secreting pituitary adenoma.

Physiologic causes, hypothyroidism, and drug-induced hyperprolactinemia should be excluded before extensive evaluation.

Pregnancy and lactation are the important physiologic causes of hyperprolactinemia. Sleep-associated hyperprolactinemia reverts to normal within an hour of awakening. Nipple stimulation and sexual orgasm also may increase PRL. Chest wall stimulation or trauma (including chest surgery and herpes zoster) invoke the reflex suckling arc with resultant hyperprolactinemia. Chronic renal failure elevates PRL by decreasing peripheral clearance. Primary hypothyroidism is associated with mild hyperprolactinemia, probably because of compensatory TRH secretion. Mutation of the PRL receptor is a rare cause of hyperprolactinemia.

Lesions of the hypothalamic-pituitary region that disrupt hypothalamic dopamine synthesis, portal vessel delivery, or lactotrope responses are associated with hyperprolactinemia. Thus, hypothalamic tumors, cysts, infiltrative disorders, and radiation-induced damage cause elevated PRL levels, usually in the range of 30–100 µg/L. Plurihormonal adenomas (including GH and ACTH tumors) may hypersecrete PRL directly. Pituitary masses, including clinically nonfunctioning pituitary tumors, may compress the pituitary stalk to cause hyperprolactinemia.

Drug-induced inhibition or disruption of dopaminergic receptor function is a common cause of hyperprolactinemia ([Table 373-5](#)). Thus, antipsychotics and antidepressants are a relatively common cause of mild hyperprolactinemia. Most patients receiving risperidone have elevated prolactin levels, sometimes exceeding 200 µg/L. Methyldopa inhibits dopamine synthesis, and [verapamil](#) blocks dopamine release, also

leading to hyperprolactinemia. Hormonal agents that induce PRL include [estrogens](#) and thyrotropin-releasing hormone (TRH).

Presentation and Diagnosis

Amenorrhea, galactorrhea, and infertility are the hallmarks of hyperprolactinemia in women. If hyperprolactinemia develops before menarche, primary amenorrhea results. More commonly, hyperprolactinemia develops later in life and leads to oligomenorrhea and ultimately to amenorrhea. If hyperprolactinemia is sustained, vertebral bone mineral density can be reduced compared with age-matched controls, particularly when it is associated with pronounced hypoestrogenemia. Galactorrhea is present in up to 80% of hyperprolactinemic women. Although usually bilateral and spontaneous, it may be unilateral or expressed only manually. Patients also may complain of decreased libido, weight gain, and mild hirsutism.

In men with hyperprolactinemia, diminished libido, infertility, and visual loss (from optic nerve compression) are the usual presenting symptoms. Gonadotropin suppression leads to reduced [testosterone](#), impotence, and oligospermia. True galactorrhea is uncommon in men with hyperprolactinemia. If the disorder is long-standing, secondary effects of hypogonadism are evident, including osteopenia, reduced muscle mass, and decreased beard growth.

The diagnosis of idiopathic hyperprolactinemia is made by exclusion of known causes of hyperprolactinemia in the setting of a normal pituitary MRI. Some of these patients may harbor small microadenomas below visible MRI sensitivity (~2 mm).

GALACTORRHEA

Galactorrhea, the inappropriate discharge of milk-containing fluid from the breast, is considered abnormal if it persists longer than 6 months after childbirth or discontinuation of breast-feeding. Postpartum galactorrhea associated with amenorrhea is a self-limiting disorder usually associated with moderately elevated PRL levels. Galactorrhea may occur spontaneously, or it may be elicited by nipple pressure. In both men and women, galactorrhea may vary in color and consistency (transparent, milky, or bloody) and arise either unilaterally or bilaterally. Mammography or ultrasound is indicated for bloody discharges (particularly from a single nipple), which may be caused by breast cancer. Galactorrhea is commonly associated with hyperprolactinemia caused by any of the [conditions](#) listed in [Table 373-5](#). Acromegaly is associated with galactorrhea in about one-third of patients. Treatment of galactorrhea usually involves managing the underlying disorder (e.g., replacing T₄ for hypothyroidism, discontinuing a medication, treating prolactinoma).

Laboratory Investigation

Basal, fasting morning PRL levels (normally <20 µg/L) should be measured to assess hypersecretion. Both false-positive and false-negative results may be encountered. In patients with markedly elevated PRL levels (>1000 µg/L), reported results may be falsely lowered because of assay artifacts; sample dilution is required to measure these high values accurately. Falsely elevated values may be caused by aggregated forms of

circulating PRL, which are usually biologically inactive (macroprolactinemia). Hypothyroidism should be excluded by measuring TSH and T₄ levels.

TREATMENT

TREATMENT

Hyperprolactinemia

Treatment of hyperprolactinemia depends on the cause of elevated PRL levels. Regardless of the etiology, however, treatment should be aimed at normalizing PRL levels to alleviate suppressive effects on gonadal function, halt galactorrhea, and preserve bone mineral density. Dopamine agonists are effective for most causes of hyperprolactinemia (see the treatment section for prolactinoma, below) regardless of the underlying cause.

If the patient is taking a medication known to cause hyperprolactinemia, the drug should be withdrawn, if possible. For psychiatric patients who require neuroleptic agents, supervised dose titration or the addition of a dopamine agonist can help restore normoprolactinemia and alleviate reproductive symptoms. However, dopamine agonists may worsen the underlying psychiatric condition, especially at high doses.

Hyperprolactinemia usually resolves after adequate thyroid hormone replacement in hypothyroid patients or after renal transplantation in patients undergoing dialysis. Resection of hypothalamic or sellar mass lesions can reverse hyperprolactinemia caused by stalk compression and reduced dopamine tone. Granulomatous infiltrates occasionally respond to glucocorticoid administration. In patients with irreversible hypothalamic damage, no treatment may be warranted. In up to 30% of patients with hyperprolactinemia—usually without a visible pituitary microadenoma—the condition may resolve spontaneously.

PROLACTINOMA

Etiology and Prevalence

Tumors arising from lactotrope cells account for about half of all functioning pituitary tumors, with a population prevalence of ~10/100,000 in men and ~30/100,000 in women. Mixed tumors that secrete combinations of GH and PRL, ACTH and PRL, and rarely TSH and PRL are also seen. These plurihormonal tumors are usually recognized by immunohistochemistry, sometimes without apparent clinical manifestations from the production of additional hormones. Microadenomas are classified as <1 cm in diameter and usually do not invade the parasellar region. Macroadenomas are >1 cm in diameter and may be locally invasive and impinge on adjacent structures. The female-to-male ratio for microprolactinomas is 20:1, whereas the sex ratio is near 1:1 for macroadenomas. Tumor size generally correlates directly with PRL concentrations; values >250 µg/L usually are associated with macroadenomas. Men tend to present with larger tumors than women, possibly because the features of male hypogonadism are less readily evident. PRL levels remain stable in most patients, reflecting the slow growth of these tumors. About 5% of microadenomas progress in the long term to macroadenomas.

Presentation and Diagnosis

Women usually present with amenorrhea, infertility, and galactorrhea. If the tumor extends outside the sella, visual field defects or other mass effects may be seen. Men often present with impotence, loss of libido, infertility, or signs of central nervous system (CNS) compression, including headaches and visual defects. Assuming that physiologic and medication-induced causes of hyperprolactinemia are excluded ([Table 373-5](#)), the diagnosis of prolactinoma is likely with a PRL level $>200 \mu\text{g/L}$. PRL levels $<100 \mu\text{g/L}$ may be caused by microadenomas, other sellar lesions that decrease dopamine inhibition, or nonneoplastic causes of hyperprolactinemia. For this reason, an MRI should be performed in all patients with hyperprolactinemia. It is important to remember that hyperprolactinemia caused secondarily by the mass effects of nonlactotrope lesions is also corrected by treatment with dopamine agonists despite failure to shrink the underlying mass. Consequently, PRL suppression by dopamine agonists does not necessarily indicate that the underlying lesion is a prolactinoma.

TREATMENT

TREATMENT

Prolactinoma

Because microadenomas rarely progress to become macroadenomas, no treatment may be needed if patients are asymptomatic and fertility is not desired; these patients should be monitored by regular serial PRL measurements and MRI scans. For symptomatic microadenomas, therapeutic goals include control of hyperprolactinemia, reduction of tumor size, restoration of menses and fertility, and resolution of galactorrhea. Dopamine agonist doses should be titrated to achieve maximal PRL suppression and restoration of reproductive function ([Fig. 373-3](#)). A normalized PRL level does not ensure reduced tumor size. However, tumor shrinkage usually is not seen in those who do not respond with lowered PRL levels. For macroadenomas, formal visual field testing should be performed before initiating dopamine agonists. MRI and visual fields should be assessed at 6- to 12-month intervals until the mass shrinks and annually thereafter until maximum size reduction has occurred.

MEDICAL

Oral dopamine agonists (cabergoline and bromocriptine) are the mainstay of therapy for patients with micro- or macroprolactinomas. Dopamine agonists suppress PRL secretion and synthesis as well as lactotrope cell proliferation. In patients with microadenomas who have achieved normoprolactinemia and significant reduction of tumor mass, the dopamine agonist may be withdrawn after 2 years. These patients should be monitored carefully for evidence of prolactinoma recurrence. About 20% of patients (especially males) are resistant to dopaminergic treatment; these adenomas may exhibit decreased D_2 dopamine receptor numbers or a postreceptor defect. D_2 receptor gene mutations in the pituitary have not been reported.

Cabergoline

An ergoline derivative, cabergoline is a long-acting dopamine agonist with high D₂ receptor affinity. The drug effectively suppresses PRL for >14 days after a single oral dose and induces prolactinoma shrinkage in most patients. Cabergoline (0.5–1.0 mg twice weekly) achieves normoprolactinemia and resumption of normal gonadal function in ~80% of patients with microadenomas; galactorrhea improves or resolves in 90% of patients. Cabergoline normalizes PRL and shrinks ~70% of macroprolactinomas. Mass effect symptoms, including headaches and visual disorders, usually improve dramatically within days after cabergoline initiation; improvement of sexual function requires several weeks of treatment but may occur before complete normalization of PRL levels. After initial control of PRL levels has been achieved, cabergoline should be reduced to the lowest effective maintenance dose. In ~5% of treated patients harboring a microadenoma, hyperprolactinemia may resolve and not recur when dopamine agonists are discontinued after long-term treatment. Cabergoline also may be effective in patients resistant to bromocriptine. Adverse effects and drug intolerance are encountered less commonly than with bromocriptine.

Bromocriptine

The ergot alkaloid bromocriptine mesylate is a dopamine receptor agonist that suppresses PRL secretion. Because it is short-acting, the drug is preferred when pregnancy is desired. In microadenomas, bromocriptine rapidly lowers serum PRL levels to normal in up to 70% of patients, decreases tumor size, and restores gonadal function. In patients with macroadenomas, PRL levels are also normalized in 70% of patients, and tumor mass shrinkage (≥50%) is achieved in most patients.

Therapy is initiated by administering a low bromocriptine dose (0.625–1.25 mg) at bedtime with a snack, followed by gradually increasing the dose. Most patients are controlled with a daily dose of <7.5 mg (2.5 mg tid).

SIDE EFFECTS

Side effects of dopamine agonists include constipation, nasal stuffiness, dry mouth, nightmares, insomnia, and vertigo; decreasing the dose usually alleviates these problems. Nausea, vomiting, and postural hypotension with faintness may occur in ~25% of patients after the initial dose. These symptoms may persist in some patients. In general, fewer side effects are reported with cabergoline. For the ~15% of patients who are intolerant of oral bromocriptine, cabergoline may be better tolerated. Intravaginal administration of bromocriptine is often efficacious in patients with intractable gastrointestinal side effects. Auditory hallucinations, delusions, and mood swings have been reported in up to 5% of patients and may be due to the dopamine agonist properties or to the lysergic acid derivative of the compounds. Rare reports of leukopenia, thrombocytopenia, pleural fibrosis, cardiac arrhythmias, and hepatitis have been described. Patients with Parkinson disease who receive at least 3 mg of cabergoline daily have been reported to be at risk for development of cardiac valve regurgitation. Studies analyzing >500 prolactinoma patients receiving recommended doses of cabergoline (up to 2 mg weekly) have shown no evidence for an increased incidence of valvular disorders. Nevertheless, because no controlled prospective studies in pituitary tumor patients are available, it is prudent to perform echocardiograms before initiating standard-dose cabergoline therapy.

Surgery

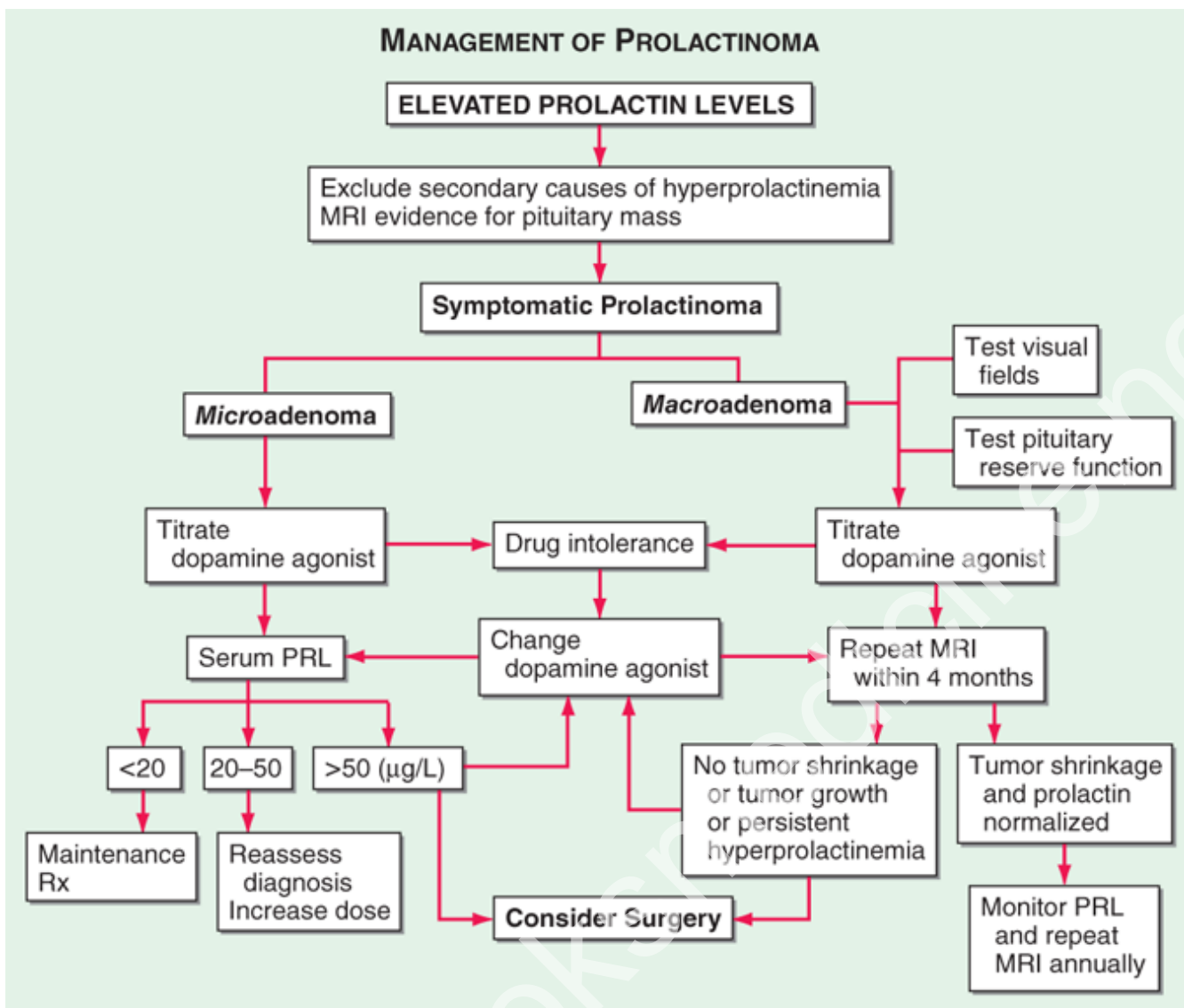
Indications for surgical adenoma debulking include dopamine resistance or intolerance and the presence of an invasive macroadenoma with compromised vision that fails to improve after drug treatment. Initial PRL normalization is achieved in ~70% of microprolactinomas after surgical resection, but only 30% of macroadenomas can be resected successfully. Follow-up studies have shown that hyperprolactinemia recurs in up to 20% of patients within the first year after surgery; long-term recurrence rates exceed 50% for macroadenomas. Radiotherapy for prolactinomas is reserved for patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists and/or surgery.

PREGNANCY

The pituitary increases in size during pregnancy, reflecting the stimulatory effects of estrogen and perhaps other growth factors on pituitary vascularity and lactotrope cell hyperplasia. About 5% of microadenomas significantly increase in size, but 15–30% of macroadenomas grow during pregnancy. Bromocriptine has been used for >30 years to restore fertility in women with hyperprolactinemia, without evidence of teratogenic effects. Nonetheless, most authorities recommend strategies to minimize fetal exposure to the drug. For women taking bromocriptine who desire pregnancy, mechanical contraception should be used through three regular menstrual cycles to allow for conception timing. When pregnancy is confirmed, bromocriptine should be discontinued and PRL levels followed serially, especially if headaches or visual symptoms occur. For women harboring macroadenomas, regular visual field testing is recommended, and the drug should be reinstituted if tumor growth is apparent. Although pituitary MRI may be safe during pregnancy, this procedure should be reserved for symptomatic patients with severe headache and/or visual field defects. Surgical decompression may be indicated if vision is threatened. Although comprehensive data support the efficacy and relative safety of bromocriptine-facilitated fertility, patients should be advised of potential unknown deleterious effects and the risk of tumor growth during pregnancy. Because cabergoline is long-acting with a high D₂-receptor affinity, it is not recommended for use in women when fertility is desired.

FIGURE 373-3

Management of prolactinoma. MRI, magnetic resonance imaging; PRL, prolactin.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

ACROMEGALY

Etiology

GH hypersecretion is usually the result of a somatotrope adenoma but may rarely be caused by extrapituitary lesions (Table 373-6). In addition to the more common GH-secreting somatotrope adenomas, mixed mammosomatotrope tumors and acidophilic stem-cell adenomas secrete both GH and PRL. In patients with acidophilic stem-cell adenomas, features of hyperprolactinemia (hypogonadism and galactorrhea) predominate over the less clinically evident signs of acromegaly. Occasionally, mixed plurihormonal tumors are encountered that also secrete ACTH, the glycoprotein hormone α subunit, or TSH in addition to GH. Patients with partially empty sellae may present with GH hypersecretion due to a small GH-secreting adenoma within the compressed rim of pituitary tissue; some of these may reflect the spontaneous necrosis of tumors that were previously larger. GH-secreting tumors rarely arise from ectopic pituitary tissue remnants in the nasopharynx or midline sinuses.

TABLE 373-6

Causes of Acromegaly

	PREVALENCE, %
Excess Growth Hormone Secretion	
Pituitary	98
Densely or sparsely granulated GH cell adenoma	60
Mixed GH cell and PRL cell adenoma	25
Mammotroph cell adenoma	10
Plurihormonal adenoma	
GH cell carcinoma or metastases	
Multiple endocrine neoplasia 1 (GH cell adenoma)	
McCune-Albright syndrome	
Ectopic sphenoid or parapharyngeal sinus pituitary adenoma	
Extrapituitary tumor	<1
Pancreatic islet cell tumor	
Lymphoma	
Excess Growth Hormone–Releasing Hormone Secretion	
Central	<1
Hypothalamic hamartoma, choristoma, ganglioglioma	
Peripheral	<1
Bronchial carcinoid, pancreatic islet cell tumor, small cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, pheochromocytoma	

Abbreviations: GH, growth hormone; PRL, prolactin.

Source: Adapted from S Melmed: N Engl J Med 355:2558, 2006.

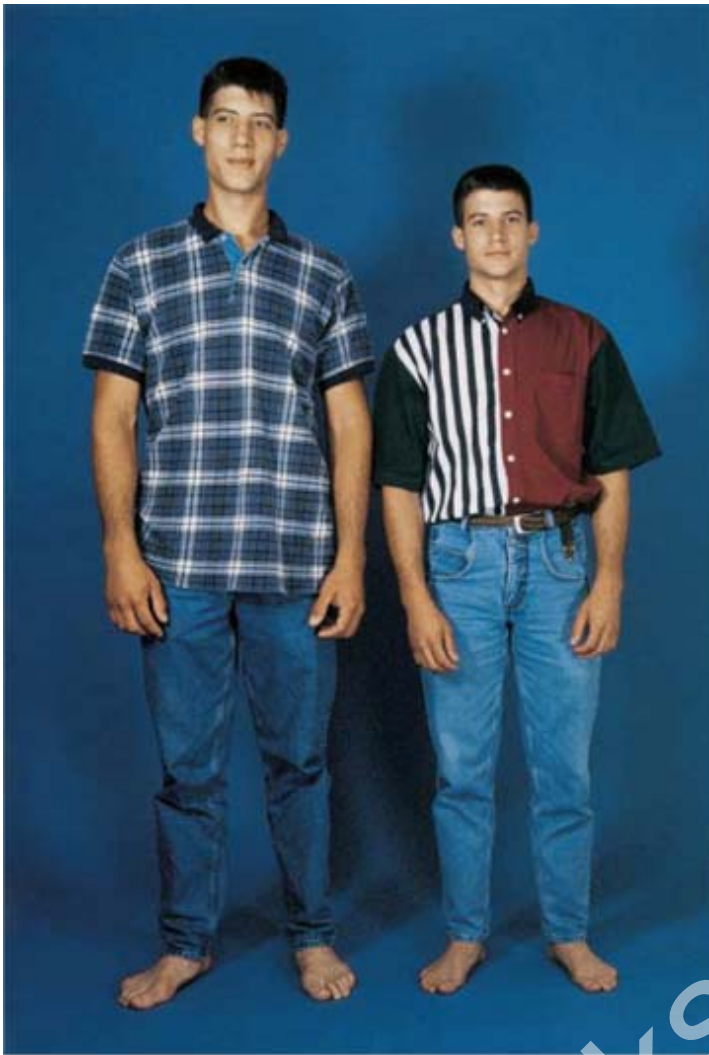
There are case reports of ectopic GH secretion by tumors of pancreatic, ovarian, lung, or hematopoietic origin. Rarely, excess GHRH production may cause acromegaly because of chronic stimulation of somatotropes. These patients present with classic features of acromegaly, elevated GH levels, pituitary enlargement on MRI, and pathologic characteristics of pituitary hyperplasia. The most common cause of GHRH-mediated acromegaly is a chest or abdominal carcinoid tumor. Although these tumors usually express positive GHRH immunoreactivity, clinical features of acromegaly are evident in only a minority of patients with carcinoid disease. Excessive GHRH also may be elaborated by hypothalamic tumors, usually choristomas or neuromas.

Presentation and Diagnosis

Protean manifestations of GH and IGF-I hypersecretion are indolent and often are not clinically diagnosed for 10 years or more. Acral bony overgrowth results in frontal bossing, increased hand and foot size, mandibular enlargement with prognathism, and widened space between the lower incisor teeth. In children and adolescents, initiation of GH hypersecretion before epiphyseal long bone closure is associated with development of pituitary gigantism (**Fig. 373-4**). Soft tissue swelling results in increased heel pad thickness, increased shoe or glove size, ring tightening, characteristic coarse facial features, and a large fleshy nose. Other commonly encountered clinical features include hyperhidrosis, a deep and hollow-sounding voice, oily skin, arthropathy, kyphosis, carpal tunnel syndrome, proximal muscle weakness and fatigue, acanthosis nigricans, and skin tags. Generalized visceromegaly occurs, including cardiomegaly, macroglossia, and thyroid gland enlargement.

FIGURE 373-4

Features of acromegaly/gigantism. A 22-year-old man with gigantism due to excess growth hormone is shown to the left of his identical twin. The increased height and prognathism (**A**) and enlarged hand (**B**) and foot (**C**) of the affected twin are apparent. Their clinical features began to diverge at the age of ~13 years. (*Reproduced from R Gagel, IE McCutcheon: N Engl J Med 324:524, 1999; with permission.*)

**B****C**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

The most significant clinical impact of GH excess occurs with respect to the cardiovascular system. Cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, and hypertension ultimately occur in most patients if untreated. Upper airway obstruction with sleep apnea occurs in >60% of patients and is associated with both soft tissue laryngeal airway obstruction and central sleep dysfunction. Diabetes mellitus develops in 25% of patients with acromegaly, and most patients are intolerant of a glucose load (as GH counteracts the action of insulin). Acromegaly is associated with an increased risk of colon polyps and mortality from colonic malignancy; polyps are diagnosed in up to one-third of patients. Overall mortality is increased about threefold and is due primarily to cardiovascular and cerebrovascular disorders and respiratory disease. Unless GH levels are controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

Laboratory Investigation

Age-matched serum IGF-I levels are elevated in acromegaly. Consequently, an IGF-I level provides a useful laboratory screening measure when clinical features raise the possibility of acromegaly. Owing to the pulsatility of GH secretion, measurement of a single random GH level is not useful for the diagnosis or

exclusion of acromegaly and does not correlate with disease severity. The diagnosis of acromegaly is confirmed by demonstrating the failure of GH suppression to $<0.4 \mu\text{g/L}$ within 1–2 h of an oral glucose load (75 g). When newer ultrasensitive GH assays are used, normal nadir GH levels are even lower ($<0.05 \mu\text{g/L}$). About 20% of patients exhibit a paradoxical GH rise after glucose. PRL should be measured, as it is elevated in ~25% of patients with acromegaly. Thyroid function, gonadotropins, and sex steroids may be attenuated because of tumor mass effects. Because most patients will undergo surgery with glucocorticoid coverage, tests of ACTH reserve in asymptomatic patients are more efficiently deferred until after surgery.

TREATMENT

TREATMENT

Acromegaly

The goal of treatment is to control GH and IGF-I hypersecretion, ablate or arrest tumor growth, ameliorate comorbidities, restore mortality rates to normal, and preserve pituitary function.

Surgical resection of GH-secreting adenomas is the initial treatment for most patients (Fig. 373-5). Somatostatin analogues are used as adjuvant treatment for preoperative shrinkage of large invasive macroadenomas, immediate relief of debilitating symptoms, and reduction of GH hypersecretion; in frail patients experiencing morbidity; and in patients who decline surgery or, when surgery fails, to achieve biochemical control. Irradiation or repeat surgery may be required for patients who cannot tolerate or do not respond to adjunctive medical therapy. The high rate of late hypopituitarism and the slow rate (5–15 years) of biochemical response are the main disadvantages of radiotherapy. Irradiation is also relatively ineffective in normalizing IGF-I levels. Stereotactic ablation of GH-secreting adenomas by gamma-knife radiotherapy is promising, but long-term results and side effects appear similar to those observed with conventional radiation. Somatostatin analogues may be required while awaiting the full benefits of radiotherapy. Systemic comorbid sequelae of acromegaly, including cardiovascular disease, diabetes, and arthritis, should be managed aggressively. Mandibular surgical repair may be indicated.

SURGERY

Transsphenoidal surgical resection by an experienced surgeon is the preferred primary treatment for both microadenomas (remission rate ~70%) and macroadenomas ($<50\%$ in remission). Soft tissue swelling improves immediately after tumor resection. GH levels return to normal within an hour, and IGF-I levels are normalized within 3–4 days. In ~10% of patients, acromegaly may recur several years after apparently successful surgery; hypopituitarism develops in up to 15% of patients after surgery.

SOMATOSTATIN ANALOGUES

Somatostatin analogues exert their therapeutic effects through SSTR2 and SSTR5 receptors, both of which are expressed by GH-secreting tumors. Octreotide acetate is an eight-amino-acid synthetic somatostatin analogue. In contrast to native somatostatin, the analogue is relatively resistant to plasma degradation. It has a 2-h serum half-life and possesses 40-fold greater potency than native somatostatin to suppress GH. Octreotide is administered by subcutaneous injection, beginning with $50 \mu\text{g tid}$; the dose can be increased

gradually up to 1500 µg/d. Octreotide suppresses integrated GH levels and normalizes IGF-I levels in ~60% of treated patients.

The long-acting somatostatin depot formulations, octreotide and lanreotide, are the preferred medical treatment for patients with acromegaly. *Octreotide LAR* is a sustained-release, long-acting formulation of octreotide incorporated into microspheres that sustain drug levels for several weeks after intramuscular injection. GH suppression occurs for as long as 6 weeks after a 30-mg intramuscular injection; long-term monthly treatment sustains GH and IGF-I suppression and also reduces pituitary tumor size in ~50% of patients. *Lanreotide Autogel*, a slow-release depot somatostatin preparation, is a cyclic somatostatin octapeptide analogue that suppresses GH and IGF-I hypersecretion after a 60-mg subcutaneous injection. Long-term (every 4–6 weeks) administration controls GH hypersecretion in about two-thirds of treated patients and improves patient compliance because of the long interval required between drug injections. Rapid relief of headache and soft tissue swelling occurs in ~75% of patients within days to weeks of somatostatin analogue initiation. Most patients report symptomatic improvement, including amelioration of headache, perspiration, obstructive apnea, and cardiac failure. For those resistant to octreotide, pasireotide, with preferential SST5 binding, has been shown to exhibit efficacy.

Side Effects

Somatostatin analogues are well tolerated in most patients. Adverse effects are short-lived and mostly relate to drug-induced suppression of gastrointestinal motility and secretion. Transient nausea, abdominal discomfort, fat malabsorption, diarrhea, and flatulence occur in one-third of patients, and these symptoms usually remit within 2 weeks. Octreotide suppresses postprandial gallbladder contractility and delays gallbladder emptying; up to 30% of patients develop long-term echogenic sludge or asymptomatic cholesterol gallstones. Other side effects include mild glucose intolerance due to transient insulin suppression, asymptomatic bradycardia, hypothyroxinemia, and local injection site discomfort. Pasireotide is associated with a higher prevalence of glucose intolerance or new-onset diabetes mellitus.

GH RECEPTOR ANTAGONIST

Pegvisomant antagonizes endogenous GH action by blocking peripheral GH binding to its receptor. Consequently, serum IGF-I levels are suppressed, reducing the deleterious effects of excess endogenous GH. Pegvisomant is administered by daily subcutaneous injection (10–20 mg) and normalizes IGF-I in ~70% of patients. GH levels, however, remain elevated as the drug does not target the pituitary adenoma. Side effects include reversible liver enzyme elevation, lipodystrophy, and injection site pain. Tumor size should be monitored by MRI.

Combined treatment with monthly somatostatin analogues and weekly or biweekly pegvisomant injections has been used effectively in resistant patients.

DOPAMINE AGONISTS

Bromocriptine and cabergoline may modestly suppress GH secretion in some patients. Very high doses of bromocriptine (≥20 mg/d) or cabergoline (0.5 mg/d) are usually required to achieve modest GH therapeutic

efficacy. Combined treatment with octreotide and cabergoline may induce additive biochemical control compared with either drug alone.

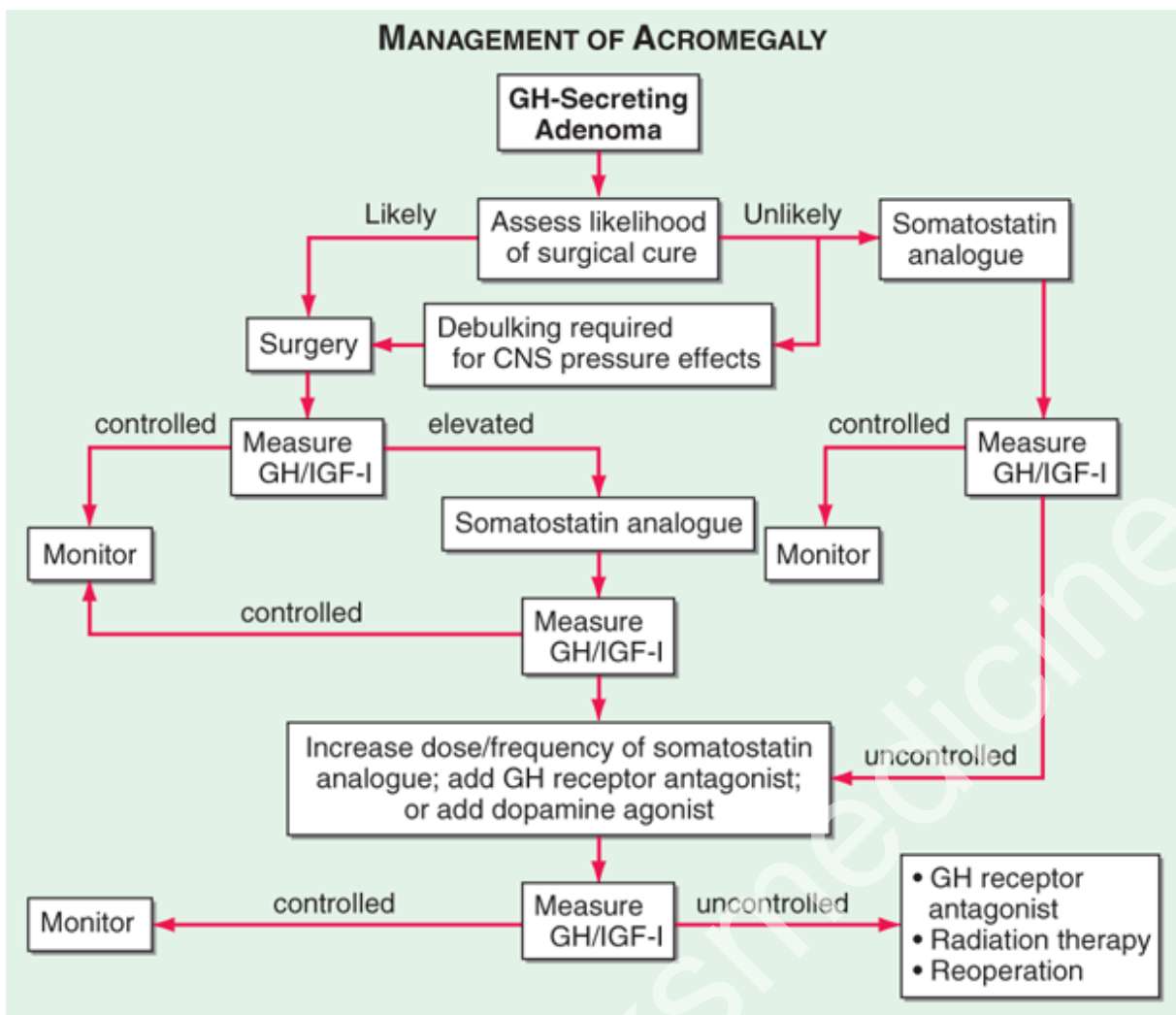
RADIATION

External radiation therapy or high-energy stereotactic techniques are used as adjuvant therapy for acromegaly. An advantage of radiation is that patient compliance with long-term treatment is not required. Tumor mass is reduced, and GH levels are attenuated over time. However, 50% of patients require at least 8 years for GH levels to be suppressed to $<5 \mu\text{g/L}$; this level of GH reduction is achieved in ~90% of patients after 18 years but represents suboptimal GH suppression. Patients may require interim medical therapy for several years before attaining maximal radiation benefits. Most patients also experience hypothalamic-pituitary damage, leading to gonadotropin, ACTH, and/or TSH deficiency within 10 years of therapy.

In summary, surgery is the preferred primary treatment for GH-secreting microadenomas (Fig. 373-5). The high frequency of GH hypersecretion after macroadenoma resection usually necessitates adjuvant or primary medical therapy for these larger tumors. Patients unable to receive or respond to unimodal medical treatment may benefit from combined treatments, or they can be offered radiation.

FIGURE 373-5

Management of acromegaly. CNS, central nervous system; IGF, insulin-like growth factor; GH, growth hormone. (Adapted from S Melmed et al: *J Clin Endocrinol Metab* 94:1509–1517, 2009; © The Endocrine Society.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

CUSHING'S DISEASE (ACTH-PRODUCING ADENOMA)

(See also [Chap. 379](#))

Etiology and Prevalence

Pituitary corticotrope adenomas (Cushing's disease) account for 70% of patients with endogenous causes of Cushing's syndrome. However, it should be emphasized that iatrogenic hypercortisolism is the most common cause of cushingoid features. Ectopic tumor ACTH production, cortisol-producing adrenal adenomas, adrenal carcinoma, and adrenal hyperplasia account for the other causes; rarely, ectopic tumor CRH production is encountered.

ACTH-producing adenomas account for ~10–15% of all pituitary tumors. Because the clinical features of Cushing's syndrome often lead to early diagnosis, most ACTH-producing pituitary tumors are relatively small microadenomas. However, macroadenomas also are seen and some ACTH-expressing adenomas are clinically silent. Cushing's disease is 5–10 times more common in women than in men. These pituitary adenomas exhibit unrestrained ACTH secretion, with resultant hypercortisolemia. However, they retain partial

suppressibility in the presence of high doses of administered glucocorticoids, providing the basis for dynamic testing to distinguish pituitary from nonpituitary causes of Cushing's syndrome.

Presentation and Diagnosis

The diagnosis of Cushing's syndrome presents two great challenges: (1) to distinguish patients with pathologic cortisol excess from those with physiologic or other disturbances of cortisol production and (2) to determine the etiology of pathologic cortisol excess.

Typical features of chronic cortisol excess include thin skin, central obesity, hypertension, plethoric moon facies, purple striae and easy bruisability, glucose intolerance or diabetes mellitus, gonadal dysfunction, osteoporosis, proximal muscle weakness, signs of hyperandrogenism (acne, hirsutism), and psychological disturbances (depression, mania, and psychoses) (Table 373-7). Hematopoietic features of hypercortisolism include leukocytosis, lymphopenia, and eosinopenia. Immune suppression includes delayed hypersensitivity and infection propensity. These protean yet commonly encountered manifestations of hypercortisolism make it challenging to decide which patients mandate formal laboratory evaluation. Certain features make pathologic causes of hypercortisolism more likely; they include characteristic central redistribution of fat, thin skin with striae and bruising, and proximal muscle weakness. In children and young females, early osteoporosis may be particularly prominent. The primary cause of death is cardiovascular disease, but life-threatening infections and risk of suicide are also increased.

TABLE 373-7

Clinical Features of Cushing's Syndrome (All Ages)

SYMPTOMS/SIGNS	FREQUENCY, %
Obesity or weight gain (>115% ideal body weight)	80
Thin skin	80
Moon facies	75
Hypertension	75
Purple skin striae	65
Hirsutism	65
Menstrual disorders (usually amenorrhea)	60
Plethora	60
Abnormal glucose tolerance	55
Impotence	55
Proximal muscle weakness	50
Truncal obesity	50
Acne	45
Bruising	45
Mental changes	45
Osteoporosis	40
Edema of lower extremities	30
Hyperpigmentation	20
Hypokalemic alkalosis	15

SYMPTOMS/SIGNS	FREQUENCY, %
Diabetes mellitus	15

Source: Adapted from MA Magiokou et al, in ME Wierman (ed): Diseases of the Pituitary. Totowa, NJ, Humana, 1997.

Rapid development of features of hypercortisolism associated with skin hyperpigmentation and severe myopathy suggests an ectopic tumor source of ACTH. Hypertension, hypokalemic alkalosis, glucose intolerance, and edema are also more pronounced in these patients. Serum potassium levels <3.3 mmol/L are evident in ~70% of patients with ectopic ACTH secretion but are seen in <10% of patients with pituitary-dependent Cushing's syndrome.

Laboratory Investigation

The diagnosis of Cushing's disease is based on laboratory documentation of endogenous hypercortisolism. Measurement of 24-h urine free cortisol (UFC) is a precise and cost-effective screening test. Alternatively, the failure to suppress plasma cortisol after an overnight 1-mg [dexamethasone](#) suppression test can be used to identify patients with hypercortisolism. As nadir levels of cortisol occur at night, elevated midnight serum or salivary samples of cortisol are suggestive of Cushing's disease. Basal plasma ACTH levels often distinguish patients with ACTH-independent (adrenal or exogenous glucocorticoid) from those with ACTH-dependent (pituitary, ectopic ACTH) Cushing's syndrome. Mean basal ACTH levels are about eightfold higher in patients with ectopic ACTH secretion than in those with pituitary ACTH-secreting adenomas. However, extensive overlap of ACTH levels in these two disorders precludes using ACTH measurements to make the distinction. Preferably, dynamic testing based on differential sensitivity to glucocorticoid feedback or ACTH stimulation in response to CRH or cortisol reduction is used to distinguish ectopic from pituitary sources of excess ACTH ([Table 373-8](#)). Very rarely, circulating CRH levels are elevated, reflecting ectopic tumor-derived secretion of CRH and often ACTH. **For further discussion of dynamic testing for Cushing's syndrome, see [Chap. 379](#).**

TABLE 373-8

Differential Diagnosis of ACTH-Dependent Cushing's Syndrome^a

	ACTH-SECRETING PITUITARY TUMOR	ECTOPIC ACTH SECRETION
Etiology	Pituitary corticotrope adenoma Plurihormonal adenoma	Bronchial, abdominal carcinoid Small cell lung cancer Thymoma
Sex	F > M	M > F
Clinical features	Slow onset	Rapid onset Pigmentation Severe myopathy
Serum potassium <3.3 µg/L	<10%	75%
24-h UFC	High	High
Basal ACTH level	Inappropriately high	Very high
Dexamethasone suppression 1 mg overnight Low-dose(0.5 mg q6h)	Cortisol >5 µg/dL	Cortisol >5 µg/dL
High-dose (2 mg q6h)	Cortisol <5 µg/dL	Cortisol >5 µg/dL
UFC >80% suppressed	Microadenomas: 90% Macroadenomas: 50%	10%
Inferior petrosal sinus sampling (IPSS)		
Basal IPSS: peripheral	>2	<2
CRH-induced IPSS: peripheral	>3	<3

^aACTH-independent causes of Cushing's syndrome are diagnosed by suppressed ACTH levels and an adrenal mass in the setting of hypercortisolism. Iatrogenic Cushing's syndrome is excluded by history.

Abbreviations: ACTH, adrenocorticotrophic hormone; CRH, corticotropin-releasing hormone; F, female; M, male; IPSS, inferior petrosal sinus sampling; UFC, urinary free cortisol.

Most ACTH-secreting pituitary tumors are <5 mm in diameter, and about half are undetectable by sensitive MRI. The high prevalence of incidental pituitary microadenomas diminishes the ability to distinguish ACTH-secreting pituitary tumors accurately from nonsecreting incidentalomas.

Inferior Petrosal Venous Sampling

Because pituitary MRI with gadolinium enhancement is insufficiently sensitive to detect small (<2 mm) pituitary ACTH-secreting adenomas, bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may be required to distinguish these lesions from ectopic ACTH-secreting tumors that may have similar clinical and biochemical characteristics. Simultaneous assessment of ACTH in each inferior petrosal vein and in the diagnosis of peripheral circulation provides a strategy for confirming and localizing pituitary ACTH production. Sampling is performed at baseline and 2, 5, and 10 min after intravenous bovine CRH (1 µg/kg) injection. An increased ratio (>2) of inferior petrosal:peripheral vein ACTH confirms pituitary Cushing's syndrome. After CRH injection, peak petrosal:peripheral ACTH ratios ≥3 confirm the presence of a pituitary ACTH-secreting tumor. The sensitivity of this test is >95%, with very rare false-positive results. False-negative results may be encountered in patients with aberrant venous drainage. Petrosal sinus catheterizations are technically difficult, and ~0.05% of patients develop neurovascular complications. The procedure should not be performed in patients with hypertension, in patients with known cerebrovascular disease, or in the presence of a well-visualized pituitary adenoma on MRI.

TREATMENT

TREATMENT

Cushing's Disease

Selective transsphenoidal resection is the treatment of choice for Cushing's disease (**Fig. 373-6**). The remission rate for this procedure is ~80% for microadenomas but <50% for macroadenomas. However, surgery is rarely successful when the adenoma is not visible on MRI. After successful tumor resection, most patients experience a postoperative period of symptomatic ACTH deficiency that may last up to 12 months. This usually requires low-dose cortisol replacement, as patients experience both steroid withdrawal symptoms and have a suppressed hypothalamic-pituitary-adrenal axis. Biochemical recurrence occurs in ~5% of patients in whom surgery was initially successful.

When initial surgery is unsuccessful, repeat surgery is sometimes indicated, particularly when a pituitary source for ACTH is well documented. In older patients, in whom issues of growth and fertility are less important, hemi- or total hypophysectomy may be necessary if a discrete pituitary adenoma is not recognized. Pituitary irradiation may be used after unsuccessful surgery, but it cures only ~15% of patients. Because the

effects of radiation are slow and only partially effective in adults, steroidogenic inhibitors are used in combination with pituitary irradiation to block adrenal effects of persistently high ACTH levels.

Pasireotide (600 or 900 µg/d subcutaneously), a somatostatin analogue with high affinity for SST5 > SST2 receptors, may control hypercortisolemia in a subset of patients with ACTH-secreting pituitary tumors when surgery is not an option or has not been successful. In clinical trials, the drug lowered plasma ACTH levels and normalized 24-h UFC levels in ~20% of patients, and resulted in up to 40% mean pituitary tumor shrinkage. Side effects include development of hyperglycemia and diabetes in up to 70% of patients, likely due to suppressed pancreatic secretion of insulin and incretins. Because patients with hypercortisolism are insulin-resistant, hyperglycemia should be rigorously managed. Other side effects are similar to those encountered for other somatostatin analogues and include transient abdominal discomfort, diarrhea, nausea, and gallstones (20% of patients). The drug requires consistent long-term administration.

Ketoconazole, an imidazole derivative antimycotic agent, inhibits several P450 enzymes and effectively lowers cortisol in most patients with Cushing's disease when administered twice daily (600–1200 mg/d). Elevated hepatic transaminases, gynecomastia, impotence, gastrointestinal upset, and edema are common side effects.

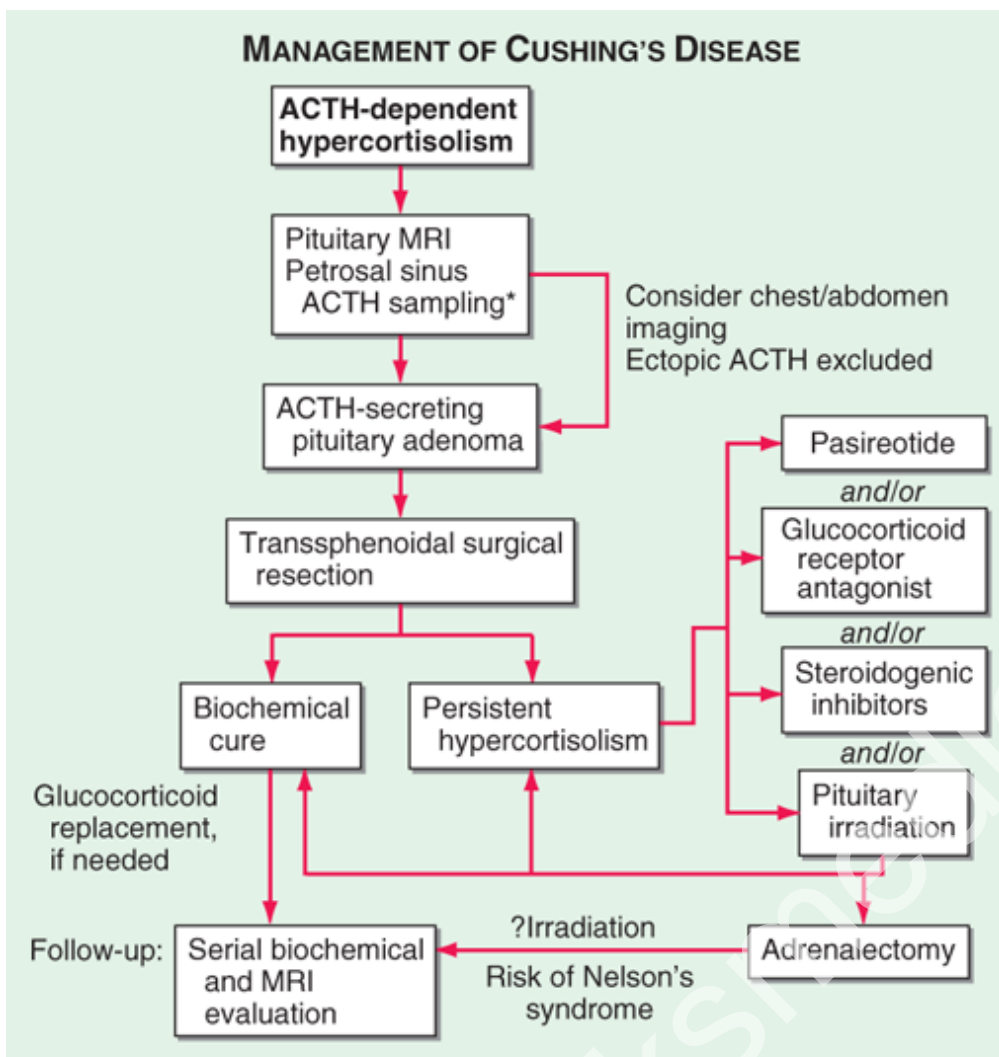
Mifepristone (300–1200 mg/d), a glucocorticoid receptor antagonist, blocks peripheral cortisol action and is approved to treat hyperglycemia in Cushing's disease. Because the drug does not target the pituitary tumor, both ACTH and cortisol levels remain elevated, thus obviating a reliable circulating biomarker. Side effects are largely due to general antagonism of other steroid hormones and include hypokalemia, endometrial hyperplasia, hypoadrenalism, and hypertension.

Metyrapone (2–4 g/d) inhibits 11β-hydroxylase activity and normalizes plasma cortisol in up to 75% of patients. Side effects include nausea and vomiting, rash, and exacerbation of acne or hirsutism. *Mitotane* (*o,p'*-DDD; 3–6 g/d orally in four divided doses) suppresses cortisol hypersecretion by inhibiting 11β-hydroxylase and cholesterol side-chain cleavage enzymes and by destroying adrenocortical cells. Side effects of mitotane include gastrointestinal symptoms, dizziness, gynecomastia, hyperlipidemia, skin rash, and hepatic enzyme elevation. It also may lead to hypoaldosteronism. Other agents include *aminoglutethimide* (250 mg tid), *trilostane* (200–1000 mg/d), *cyproheptadine* (24 mg/d), and IV *etomidate* (0.3 mg/kg/h). Glucocorticoid insufficiency is a potential side effect of agents used to block steroidogenesis.

The use of steroidogenic inhibitors has decreased the need for bilateral adrenalectomy. Surgical removal of both adrenal glands corrects hypercortisolism but may be associated with significant morbidity rates and necessitates permanent glucocorticoid and mineralocorticoid replacement. Adrenalectomy in the setting of residual corticotrope adenoma tissue predisposes to the development of *Nelson's syndrome*, a disorder characterized by rapid pituitary tumor enlargement and increased pigmentation secondary to high ACTH levels. Prophylactic radiation therapy may be indicated to prevent the development of Nelson's syndrome after adrenalectomy.

FIGURE 373-6

Management of Cushing's disease. ACTH, adrenocorticotropin hormone; MRI, magnetic resonance imaging; *, Not usually required.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

NONFUNCTIONING AND GONADOTROPIN-PRODUCING PITUITARY ADENOMAS

Etiology and Prevalence

Nonfunctioning pituitary adenomas include those that secrete little or no pituitary hormones as well as tumors that produce too little hormone to result in recognizable clinical features. They are the most common type of pituitary adenoma and are usually macroadenomas at the time of diagnosis because clinical features are not apparent until tumor mass effects occur. Based on immunohistochemistry, most clinically nonfunctioning adenomas can be shown to originate from gonadotrope cells. These tumors typically produce small amounts of intact gonadotropins (usually FSH) as well as uncombined α , LH β , and FSH β subunits. Tumor secretion may lead to elevated α and FSH β subunits and, rarely, to increased LH β subunit levels. Some adenomas express α subunits without FSH or LH. TRH administration often induces an atypical increase of tumor-derived gonadotropins or subunits.

Presentation and Diagnosis

Clinically nonfunctioning tumors often present with optic chiasm pressure and other symptoms of local expansion or may be incidentally discovered on an MRI performed for another indication (incidentaloma). Rarely, menstrual disturbances or ovarian hyperstimulation occur in women with large tumors that produce FSH and LH. More commonly, adenoma compression of the pituitary stalk or surrounding pituitary tissue leads to attenuated LH and features of hypogonadism. PRL levels are usually slightly increased, also because of stalk compression. It is important to distinguish this circumstance from true prolactinomas, as nonfunctioning tumors do not shrink in response to treatment with dopamine agonists.

Laboratory Investigation

The goal of laboratory testing in clinically nonfunctioning tumors is to classify the type of the tumor, identify hormonal markers of tumor activity, and detect possible hypopituitarism. Free α subunit levels may be elevated in 10–15% of patients with nonfunctioning tumors. In female patients, peri- or postmenopausal basal FSH concentrations are difficult to distinguish from tumor-derived FSH elevation. Premenopausal women have cycling FSH levels, also preventing clear-cut diagnostic distinction from tumor-derived FSH. In men, gonadotropin-secreting tumors may be diagnosed because of slightly increased gonadotropins (FSH > LH) in the setting of a pituitary mass. Testosterone levels are usually low despite the normal or increased LH level, perhaps reflecting reduced LH bioactivity or the loss of normal LH pulsatility. Because this pattern of hormone test results is also seen in primary gonadal failure and, to some extent, with aging ([Chap. 384](#)), the finding of increased gonadotropins alone is insufficient for the diagnosis of a gonadotropin-secreting tumor. In the majority of patients with gonadotrope adenomas, TRH administration stimulates LH β subunit secretion; this response is not seen in normal individuals. GnRH testing, however, is not helpful for making the diagnosis. For nonfunctioning and gonadotropin-secreting tumors, the diagnosis usually rests on immunohistochemical analyses of surgically resected tumor tissue, as the mass effects of these tumors usually necessitate resection.

Although acromegaly or Cushing's syndrome usually presents with unique clinical features, clinically inapparent (silent) somatotrope or corticotrope adenomas may only be diagnosed by immunostaining of resected tumor tissue. If PRL levels are <100 $\mu\text{g/L}$ in a patient harboring a pituitary mass, a nonfunctioning adenoma causing pituitary stalk compression should be considered.

TREATMENT

TREATMENT

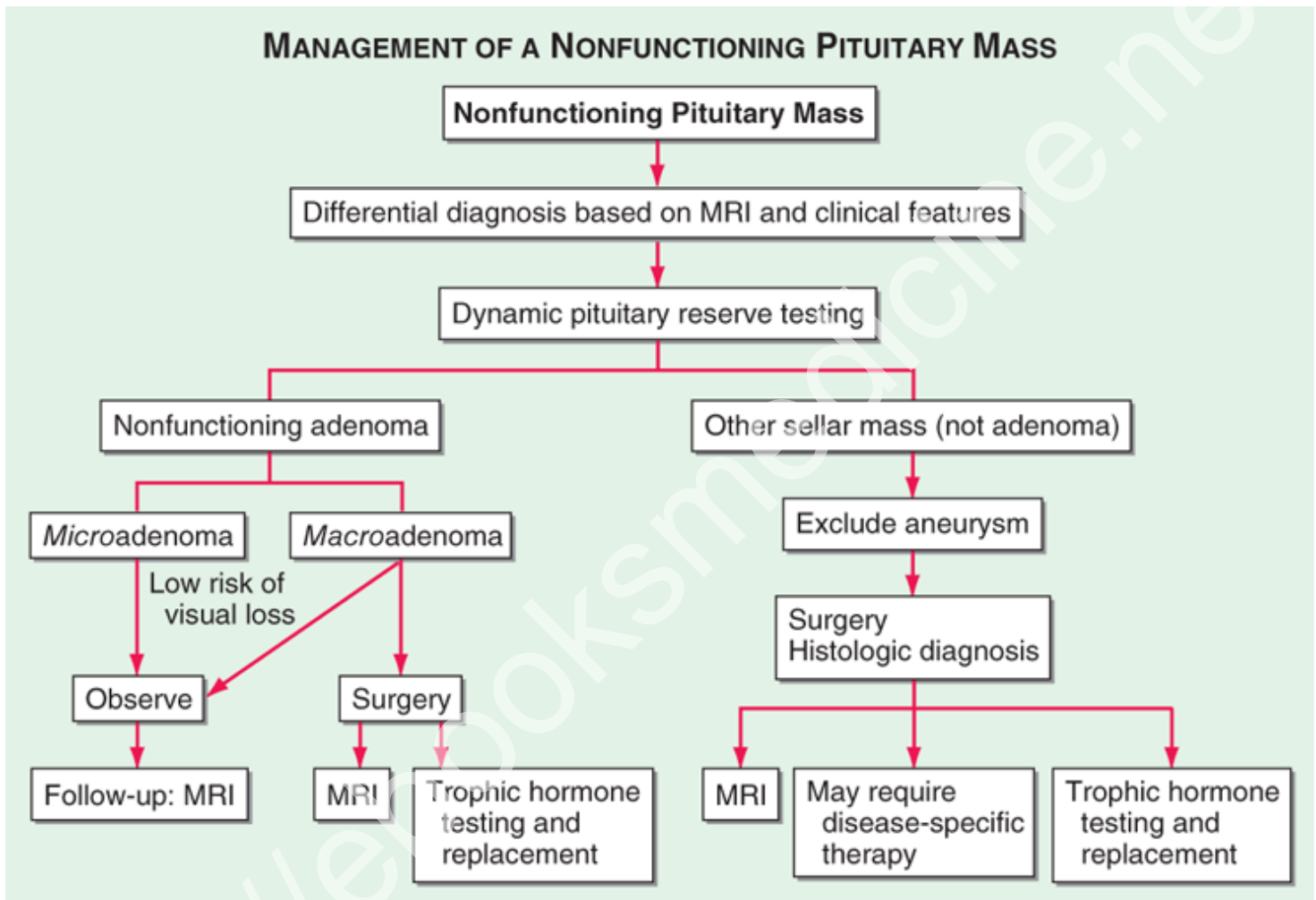
Nonfunctioning and Gonadotropin-Producing Pituitary Adenomas

Asymptomatic small nonfunctioning microadenomas with no threat to vision may be followed with regular MRI and visual field testing without immediate intervention. However, for macroadenomas, transsphenoidal surgery is indicated to reduce tumor size and relieve mass effects ([Fig. 373-7](#)). Although it is not usually possible to remove all adenoma tissue surgically, vision improves in 70% of patients with preoperative visual field defects. Preexisting hypopituitarism that results from tumor mass effects may improve or resolve completely. Beginning ~6 months postoperatively, MRI scans should be performed yearly to detect tumor regrowth. Within 5–6 years after successful surgical resection, ~15% of nonfunctioning tumors recur. When

substantial tumor remains after transsphenoidal surgery, adjuvant radiotherapy may be indicated to prevent tumor regrowth. Radiotherapy may be deferred if no postoperative residual mass is evident. Nonfunctioning pituitary tumors respond poorly to dopamine agonist treatment and somatostatin analogues are largely ineffective for shrinking these tumors. The selective GnRH antagonist Nal-Glu GnRH suppresses FSH hypersecretion but has no effect on adenoma size.

FIGURE 373-7

Management of a nonfunctioning pituitary mass. MRI, magnetic resonance imaging.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

TSH-SECRETING ADENOMAS

TSH-producing macroadenomas are very rare but are often large and locally invasive when they occur. Patients usually present with thyroid goiter and hyperthyroidism, reflecting overproduction of TSH. Diagnosis is based on demonstrating elevated serum-free T₄ levels, inappropriately normal or high TSH secretion, and MRI evidence of a pituitary adenoma. Elevated uncombined α subunits are seen in many patients.

It is important to exclude other causes of inappropriate TSH secretion, such as resistance to thyroid hormone, an autosomal dominant disorder caused by mutations in the thyroid hormone β receptor (Chap. 375). The

presence of a pituitary mass and elevated β subunit levels are suggestive of a TSH-secreting tumor. Dysalbuminemic hyperthyroxinemia syndromes, caused by mutations in serum thyroid hormone binding proteins, are also characterized by elevated thyroid hormone levels, but with normal rather than suppressed TSH levels. Moreover, free thyroid hormone levels are normal in these disorders, most of which are familial.

TREATMENT

TREATMENT

TSH-Secreting Adenomas

The initial therapeutic approach is to remove or debulk the tumor mass surgically, usually using a transsphenoidal approach. Total resection is not often achieved as most of these adenomas are large and locally invasive. Normal circulating thyroid hormone levels are achieved in about two-thirds of patients after surgery. Thyroid ablation or antithyroid drugs (methimazole and propylthiouracil) can be used to reduce thyroid hormone levels. Somatostatin analogue treatment effectively normalizes TSH and α subunit hypersecretion, shrinks the tumor mass in 50% of patients, and improves visual fields in 75% of patients; euthyroidism is restored in most patients. Because somatostatin analogues markedly suppress TSH, biochemical hypothyroidism often requires concomitant thyroid hormone replacement, which may also further control tumor growth.

FURTHER READING

Di Ieva A et al: Aggressive pituitary adenomas—diagnosis and emerging treatments. *Nat Rev Endocrinol* 10:423, 2014. [[PubMed: 24821329](#)]

Famini P, Maya MM, Melmed S: Pituitary magnetic resonance imaging for sellar and parasellar masses: Ten-year experience in 2598 patients. *J Clin Endocrinol Metab* 96:1633, 2011. [[PubMed: 21470998](#)]

Freda PU et al: Pituitary incidentaloma: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 96:894, 2011. [[PubMed: 21474686](#)]

Gittleman H et al: Descriptive epidemiology of pituitary tumors in the United States, 2004–2009. *J Neurosurg* 121:527, 2014. [[PubMed: 24926650](#)]

Hannah-Shmouni F, Trivellin G, Stratakis CA: Genetics of gigantism and acromegaly. *Growth Horm IGF Res* 30–31:37–41, 2016. [[PubMed: 27657986](#)]

Melmed S: Medical progress: acromegaly. *N Engl J Med* 355:2558, 2006. [[PubMed: 17167139](#)]

Melmed S: Pathogenesis of pituitary tumors. *Nat Rev Endocrinol* 7:257, 2011. [[PubMed: 21423242](#)]

Melmed S et al: Diagnosis and treatment of hyperprolactinemia: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 96:273, 2011. [[PubMed: 21296991](#)]

Nieman LK: Cushing's syndrome: Update on signs, symptoms and biochemical screening. Eur J Endocrinol 173:M33, 2015.

Ntali G et al: Clinical review: Functioning gonadotroph adenomas. J Clin Endocrinol Metab 99:4423, 2014. [[PubMed: 25166722](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 374: Disorders of the Neurohypophysis

Gary L. Robertson

INTRODUCTION

The neurohypophysis, or posterior pituitary, is formed by axons that originate in large cell bodies in the supraoptic and paraventricular nuclei of the hypothalamus. It produces two hormones: (1) arginine vasopressin (AVP), also known as antidiuretic hormone, and (2) oxytocin. AVP acts on the renal tubules to reduce water loss by concentrating the urine. Oxytocin stimulates postpartum milk letdown in response to suckling. A deficiency of AVP secretion or action causes diabetes insipidus (DI), a syndrome characterized by the production of large amounts of dilute urine. Excessive or inappropriate AVP production impairs urinary water excretion and predisposes to hyponatremia if water intake is not reduced in parallel with urine output.

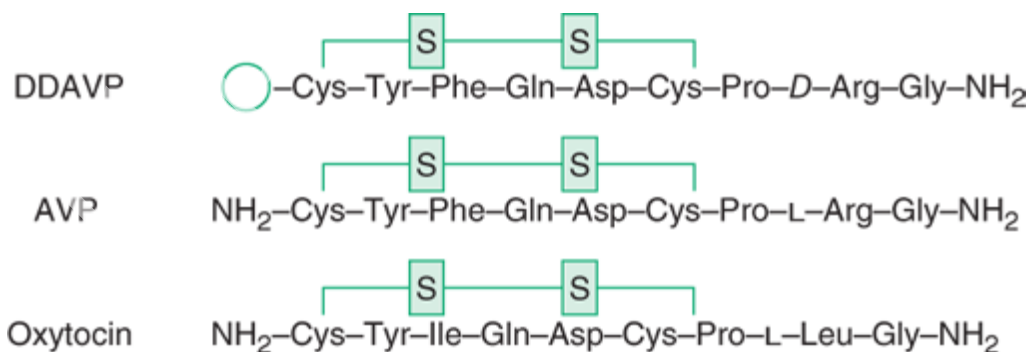
VASOPRESSIN

SYNTHESIS AND SECRETION

AVP is a nonapeptide composed of a six-member disulfide ring and a tripeptide tail (**Fig. 374-1**). It is synthesized via a polypeptide precursor that includes AVP, neurophysin, and copeptin, all encoded by a single gene on chromosome 20. After preliminary processing and folding, the precursor is packaged in neurosecretory vesicles, where it is transported down the axon; further processed to AVP, neurophysin, and copeptin; and stored in neurosecretory vesicles until released by exocytosis into peripheral blood.

FIGURE 374-1

Primary structures of arginine vasopressin (AVP), oxytocin, and desmopressin (DDAVP).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

AVP secretion is regulated primarily by the “effective” osmotic pressure of body fluids. This control is mediated by specialized hypothalamic cells known as *osmoreceptors*, which are extremely sensitive to small changes in the plasma concentration of sodium and its anions but normally are insensitive to other solutes such as [urea](#) and glucose. The osmoreceptors appear to include inhibitory as well as stimulatory components that function in concert to form a threshold, or set point, control system. Below this threshold, plasma AVP is suppressed to levels that permit the development of a maximum water diuresis. Above it, plasma AVP rises steeply in direct proportion to plasma osmolality, quickly reaching levels sufficient to effect a maximum antidiuresis. The absolute levels of plasma osmolality/sodium at which minimally and maximally effective levels of plasma AVP occur, vary appreciably from person to person, apparently due to genetic influences on the set and sensitivity of the system. However, the average threshold, or set point, for AVP release corresponds to a plasma osmolality or sodium of about 280 mosmol/L or 135 meq/L, respectively; levels only 2–4% higher normally result in maximum antidiuresis.

Although it is relatively stable in a healthy adult, the set point of the osmoregulatory system can be lowered by pregnancy, the menstrual cycle, estrogen, and relatively large, acute reductions in blood pressure or volume. Those reductions are mediated largely by neuronal afferents that originate in transmural pressure receptors of the heart and large arteries and project via the vagus and glossopharyngeal nerves to the brainstem, from which postsynaptic projections ascend to the hypothalamus. These pathways maintain a tonic inhibitory tone that decreases when blood volume or pressure falls by >10–20%. This baroregulatory system is probably of minor importance in the physiology of AVP secretion because the hemodynamic changes required to affect it usually do not occur during normal activities. However, the baroregulatory system undoubtedly plays an important role in AVP secretion in patients with disorders that produce large, acute disturbances of hemodynamic function. AVP secretion also can be stimulated by nausea, acute hypoglycemia, glucocorticoid deficiency, [smoking](#), and, possibly, hyperangiotensinemia. The emetic stimuli are extremely potent since they typically elicit immediate, 50- to 100-fold increases in plasma AVP even when the nausea is transient and is not associated with vomiting or other symptoms. They appear to act via the emetic center in the medulla and can be blocked completely by treatment with antiemetics such as fluphenazine. There is no evidence that pain or other noxious stresses have any effect on AVP unless they elicit a vasovagal reaction with its associated nausea and hypotension.

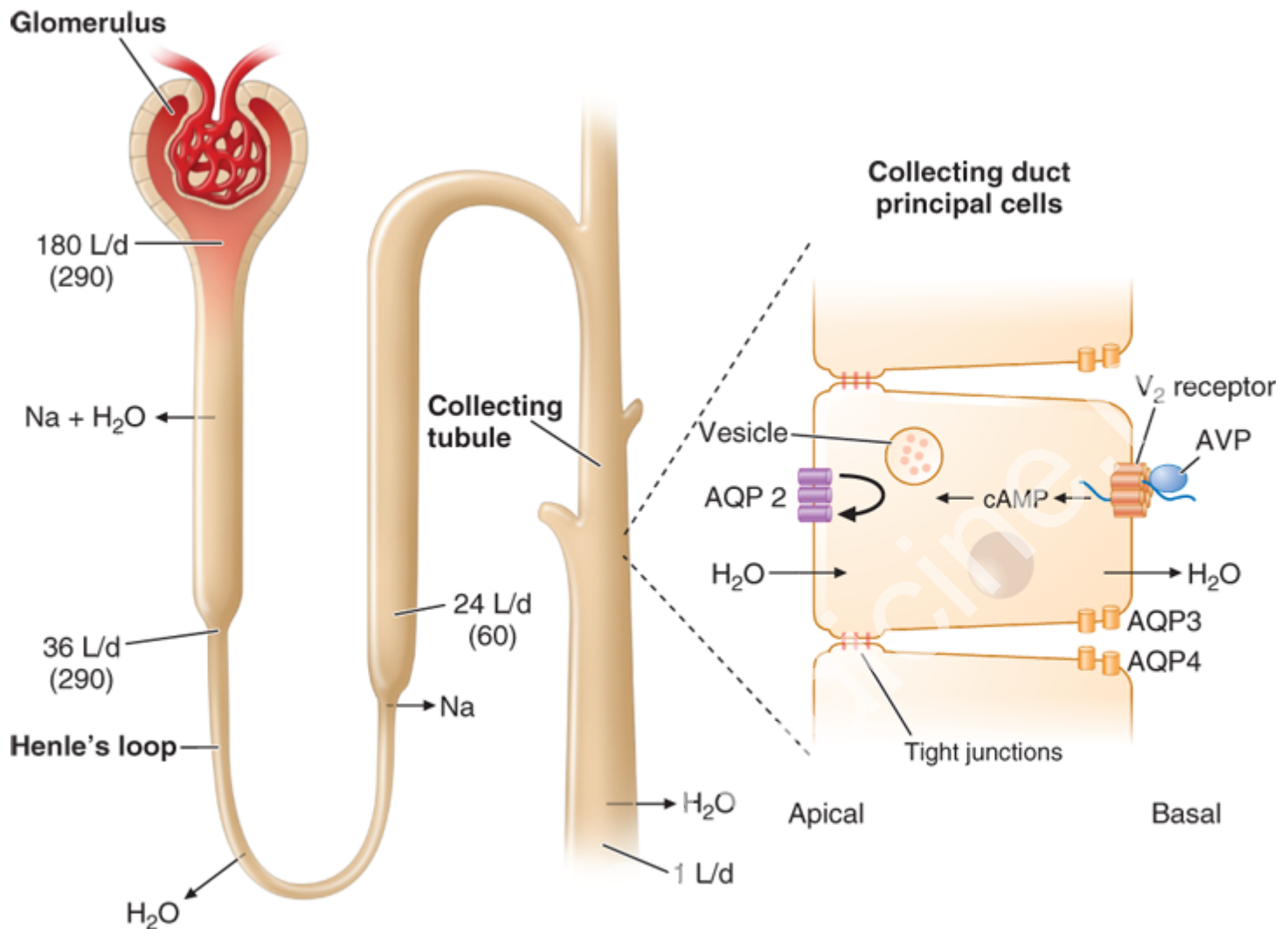
ACTION

The most important, if not the only, physiologic action of AVP is to reduce water excretion by promoting concentration of urine. This antidiuretic effect is achieved by increasing the hydroosmotic permeability of cells that line the distal tubule and medullary collecting ducts of the kidney ([Fig. 374-2](#)). In the absence of AVP, these cells are impermeable to water and reabsorb little, if any, of the relatively large volume of dilute filtrate that enters from the proximal nephron. The lack of reabsorption results in the excretion of very large volumes (as much as 0.2 mL/kg per min) of maximally dilute urine (specific gravity and osmolality ~1.000 and 50 mosmol/L, respectively), a condition known as *water diuresis*. In the presence of AVP, these cells become selectively permeable to water, allowing the water to diffuse back down the osmotic gradient created by the hypertonic renal medulla. As a result, the dilute fluid passing through the tubules is concentrated and the

rate of urine flow decreases. The magnitude of this effect varies in direct proportion to the plasma AVP concentration and the rate of solute excretion. At maximum levels of AVP and normal rates of solute excretion, it approximates a urine flow rate as low as 0.35 mL/min and a urine osmolarity as high as 1200 mosmol/L. This effect is reduced by a solute diuresis such as glucosuria in diabetes mellitus.

FIGURE 374-2

Antidiuretic effect of arginine vasopressin (AVP) in the regulation of urine volume. In a typical 70-kg adult, the kidney filters ~180 L/d of plasma. Of this, ~144 L (80%) is reabsorbed isosmotically in the proximal tubule and another 8 L (4–5%) is reabsorbed without solute in the descending limb of Henle's loop. The remainder is diluted to an osmolarity of ~60 mmol/kg by selective reabsorption of sodium and chloride in the ascending limb. In the absence of AVP, the urine issuing from the loop passes largely unmodified through the distal tubules and collecting ducts, resulting in a maximum water diuresis. In the presence of AVP, solute-free water is reabsorbed osmotically through the principal cells of the collecting ducts, resulting in the excretion of a much smaller volume of concentrated urine. This antidiuretic effect is mediated via a G protein–coupled V_2 receptor that increases intracellular cyclic AMP, thereby inducing translocation of aquaporin 2 (AQP 2) water channels into the apical membrane. The resultant increase in permeability permits an influx of water that diffuses out of the cell through AQP 3 and AQP 4 water channels on the basal-lateral surface. The net rate of flux across the cell is determined by the number of AQP 2 water channels in the apical membrane and the strength of the osmotic gradient between tubular fluid and the renal medulla. Tight junctions on the lateral surface of the cells serve to prevent unregulated water flow. The V_2 receptors and aquaporin 2 are encoded by genes on chromosome Xq28 and 12q13, respectively.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

At high concentrations, AVP also causes contraction of smooth muscle in blood vessels in the skin and gastrointestinal tract, induces glycogenolysis in the liver, and potentiates adrenocorticotrophic hormone (ACTH) release by corticotropin-releasing factor. These effects are mediated by V_{1a} or V_{1b} receptors that are coupled to phospholipase C. Their role, if any, in human physiology/pathophysiology is uncertain.

METABOLISM

AVP distributes rapidly into a space roughly equal to the extracellular fluid volume. It is cleared irreversibly with a half-life ($t_{1/2}$) of 10–30 min. Most AVP clearance is due to degradation in the liver and kidneys. During pregnancy, the metabolic clearance of AVP is increased three- to fourfold due to placental production of an N-terminal peptidase.

THIRST

Because AVP cannot reduce water loss below a certain minimum level obligated by urinary solute load and evaporation from skin and lungs, a mechanism for ensuring adequate intake is essential for preventing

dehydration. This vital function is performed by the thirst mechanism. Like AVP, thirst is regulated primarily by an osmostat that is situated in the anteromedial hypothalamus and is able to detect very small changes in the plasma concentration of sodium and its anions. The thirst osmostat appears to be “set” about 3% higher than the AVP osmostat. This arrangement ensures that thirst, polydipsia, and dilution of body fluids do not occur until plasma osmolality/sodium starts to exceed the defensive capacity of the antidiuretic mechanism.

OXYTOCIN

Oxytocin is also a nonapeptide that differs from AVP only at positions 3 and 8 (Fig. 374-1). However, it has relatively little antidiuretic effect and seems to act mainly on mammary ducts to facilitate milk letdown during nursing. It also may help initiate or facilitate labor by stimulating contraction of uterine smooth muscle, but it is not clear if this action is physiologic or necessary for normal delivery.

DEFICIENCIES OF AVP SECRETION AND ACTION

DIABETES INSIPIDUS

Clinical Characteristics

A decrease of 75% or more in the secretion or action of AVP usually results in DI, a syndrome characterized by the production of abnormally large volumes of dilute urine. The 24-h urine volume exceeds 40 mL/kg body weight, and the osmolality is <300 mosmol/L. The polyuria produces symptoms of urinary frequency, enuresis, and/or nocturia, which may disturb sleep and cause mild daytime fatigue or somnolence. It also results in a slight rise in plasma osmolality that stimulates thirst and a commensurate increase in fluid intake (polydipsia). Overt clinical signs of dehydration are uncommon unless thirst and/or the compensatory increase of fluid intake are also impaired.

Etiology



A primary deficiency of AVP secretion usually results from agenesis or irreversible destruction of the neurohypophysis. It is referred to variously as *neurohypophyseal DI*, *neurogenic DI*, *pituitary DI*, *cranial DI*, or *central DI*. It can be caused by a variety of congenital, acquired, or genetic disorders, but in about one-half of all adult patients, it is idiopathic (Table 374-1). Pituitary DI caused by surgery in or around the neurohypophysis usually appears within 24 h. After a few days, it may transition to a 2- to 3-week period of inappropriate antidiuresis, after which the DI may or may not recur permanently. Five genetic forms of pituitary DI are now known. By far, the most common is transmitted in an autosomal dominant mode and is caused by diverse mutations in the coding region of one allele of the AVP–neurophysin II (or *AVP-NPII*) gene. All the mutations alter one or more amino acids known to be critical for correct processing and/or folding of the prohormone, thus interfering with its trafficking through the endoplasmic reticulum. The misfolded mutant precursor accumulates and interferes with production of AVP by the normal allele, eventually

destroying the magnocellular neurons in which it is produced. The AVP deficiency and DI are usually not present at birth but develop gradually over a period of several months to years, progressing from partial to severe at different rates depending on the mutation. Once established, the deficiency of AVP is permanent, but for unknown reasons, the DI occasionally improves or remits spontaneously in late middle age. The parvocellular neurons that make AVP and the magnocellular neurons that make oxytocin appear to be unaffected. There are also rare autosomal recessive forms of pituitary DI. One is due to an inactivating mutation in the AVP portion of the gene; another is due to a large deletion involving the majority of the AVP gene and regulatory sequences in the intergenic region. A third form is caused by mutations of the *WFS 1* gene responsible for Wolfram's syndrome (DI, diabetes mellitus, optic atrophy, and neural deafness [DIDMOAD]). An X-linked recessive form linked to a region on Xq28 has also been described but the causative gene has not yet been identified.

TABLE 374-1

Causes of Diabetes Insipidus

Pituitary diabetes insipidus Acquired Head trauma (closed and penetrating) including pituitary surgery Neoplasms Primary Craniopharyngioma Pituitary adenoma (suprasellar) Dysgerminoma Meningioma Metastatic (lung, breast) Hematologic (lymphoma, leukemia) Granulomas Sarcoidosis Histiocytosis Xanthoma disseminatum Infectious Chronic meningitis Viral encephalitis Toxoplasmosis Inflammatory Lymphocytic infundibuloneurohypophysitis Granulomatosis with polyangiitis (Wegener's) Lupus erythematosus Scleroderma Chemical toxins Tetrodotoxin Snake venom Vascular Sheehan's syndrome Aneurysm (internal carotid) Aortocoronary bypass Hypoxic encephalopathy Idiopathic Congenital malformations Septo-optic dysplasia Midline craniofacial defects	Gestational diabetes insipidus Pregnancy (second and third trimesters) Nephrogenic diabetes insipidus Acquired Drugs Lithium Demeclocycline Methoxyflurane Amphotericin B Aminoglycosides Cisplatin Rifampin Foscarnet Metabolic Hypercalcemia, hypercalciuria Hypokalemia Obstruction (ureter or urethra) Vascular Sickle cell disease and trait Ischemia (acute tubular necrosis) Granulomas Sarcoidosis Neoplasms Sarcoma Infiltration Amyloidosis Idiopathic Genetic X-linked recessive (<i>AVP receptor-2</i> gene) Autosomal recessive (<i>AQP2</i> gene) Autosomal dominant (<i>AQP2</i> gene) Primary polydipsia Acquired Psychogenic Schizophrenia Obsessive compulsive disorder
--	--

Holoprosencephaly	Dipsogenic (abnormal thirst)
Hypogenesis, ectopia of pituitary	Granulomas (sarcoidosis)
Genetic	Infectious (tuberculous meningitis)
Autosomal dominant	Head trauma (closed and penetrating)
(<i>AVP-neurophysin</i> gene)	Demyelination (multiple sclerosis)
Autosomal recessive	Drugs
Type A (<i>AVP-neurophysin</i> gene)	Idiopathic
Type B (<i>AVP-neurophysin</i> gene)	Iatrogenic
Type C (<i>Wolfram's [4p-WFS 1]</i> gene)	
X-linked recessive (Xq28)	

A primary deficiency of plasma AVP also can result from increased metabolism by an N-terminal aminopeptidase produced by the placenta. It is referred to as *gestational DI* because the signs and symptoms manifest during pregnancy and usually remit several weeks after delivery.

Secondary deficiencies of AVP secretion result from inhibition by excessive intake of fluids. They are referred to as *primary polydipsia* and can be divided into three subcategories. One of them, *dipsogenic DI*, is characterized by inappropriate thirst caused by a reduction in the set of the osmoregulatory mechanism. It sometimes occurs in association with multifocal diseases of the brain such as neurosarcoid, tuberculous meningitis, and multiple sclerosis but is often idiopathic. The second subtype, *psychogenic polydipsia*, is not associated with thirst, and the polydipsia seems to be a feature of psychosis or obsessive compulsive disorder. The third subtype, *iatrogenic polydipsia*, results from recommendations to increase fluid intake for its presumed health benefits.

Primary deficiencies in the antidiuretic action of AVP result in *nephrogenic DI*. The causes can be genetic, acquired, or drug induced (Table 374-1). The most common genetic form is transmitted in a semirecessive X-linked manner. It is caused by mutations in the coding region of the V_2 receptor gene that impair trafficking and/or ligand binding of the mutant receptor. There are also autosomal recessive or dominant forms of nephrogenic DI. They are caused by *AQP2* gene mutations that result in complete or partial defects in trafficking and function of the water channels that mediate antidiuresis in the distal and collecting tubules of the kidney.

Secondary deficiencies in the antidiuretic response to AVP result from polyuria per se. They are caused by washout of the medullary concentration gradient and/or suppression of aquaporin function. They usually resolve 24–48 h after the polyuria is corrected but can complicate interpretation of some acute tests used for differential diagnosis.

Pathophysiology

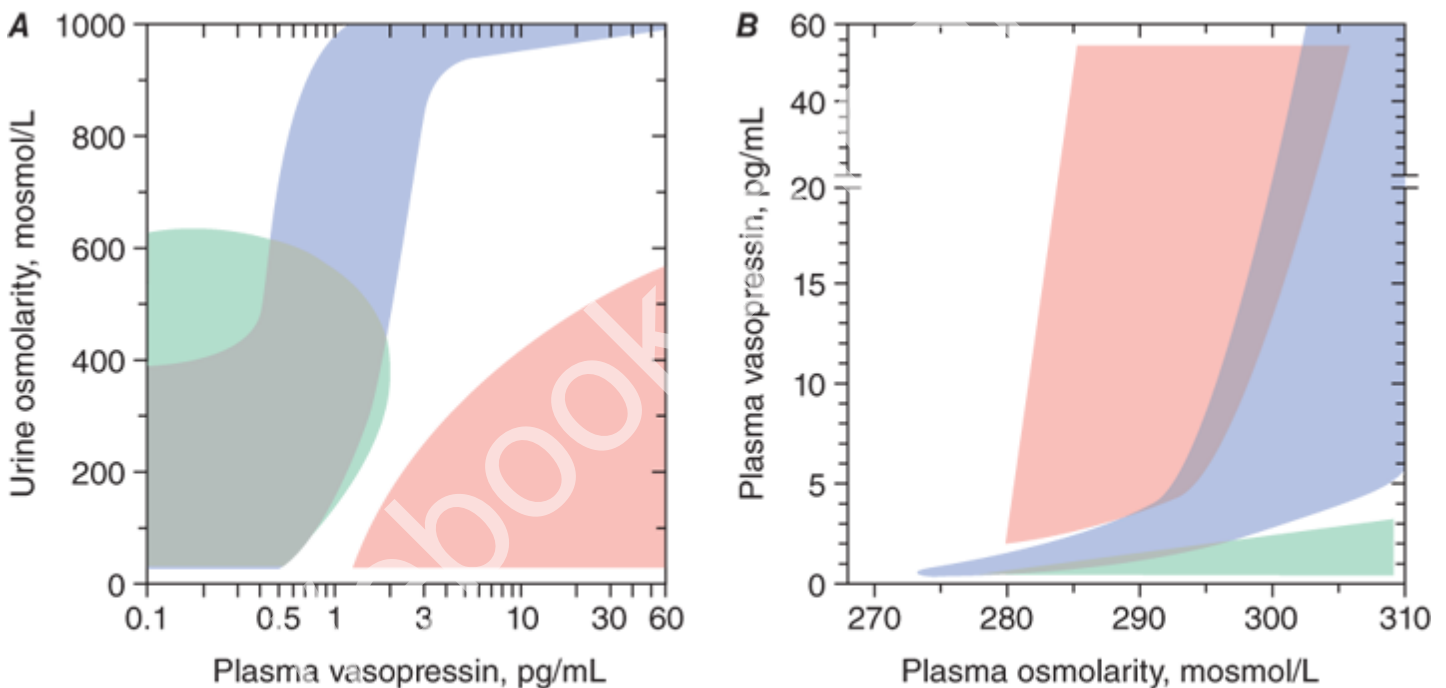
In pituitary, gestational, or nephrogenic DI, the polyuria results in a small (1–2%) decrease in body water and a commensurate increase in plasma osmolality and sodium that stimulates thirst and a compensatory

increase in water intake. As a result, *hypernatremia and other overt physical or laboratory signs of dehydration do not develop unless the patient also has a defect in thirst or fails to increase fluid intake for some other reason.*

In pituitary and nephrogenic DI, the severity of the defect in AVP secretion or action varies significantly from patient to patient. In some, the defect is so severe that it cannot be overcome by even an intense stimulus such as nausea or severe dehydration. In others, the defect in AVP secretion or action is incomplete, and a modest stimulus such as a few hours of fluid deprivation, smoking, or a vasovagal reaction can raise urine osmolarity as high as 800 mosmol/L. However, even when the defects are partial, the relation of urine osmolarity to plasma AVP in patients with nephrogenic DI (Fig. 374-3A) or of plasma AVP to plasma osmolarity and sodium in patients with pituitary DI (Fig. 374-3B) is subnormal.

FIGURE 374-3

Relationship of plasma AVP to urine osmolarity (A) and plasma osmolarity (B) before and during fluid deprivation–hypertonic saline infusion test in patients who are normal or have primary polydipsia (blue zones), pituitary diabetes insipidus (green zones), or nephrogenic diabetes insipidus (pink zones).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

In primary polydipsia, the pathogenesis of the polydipsia and polyuria is the reverse of that in pituitary, nephrogenic, and gestational DI. In primary polydipsia, an abnormality in cognition or thirst causes excessive intake of fluids and an increase in body water that reduces plasma osmolarity/sodium, AVP secretion, and urinary concentration. Dilution of the urine, in turn, results in a compensatory increase in urinary free-water excretion that usually offsets the increase in intake and stabilizes plasma osmolarity/sodium at a level only 1–2% below basal. Thus, hyponatremia or clinically appreciable overhydration is uncommon unless the polydipsia is very severe or the compensatory water diuresis is impaired by a drug or disease that stimulates or mimics the antidiuretic effect of endogenous AVP. A rise in plasma osmolarity and sodium produced by

fluid deprivation or hypertonic saline infusion increases plasma AVP normally. However, the resultant increase in urine concentration is often subnormal because polyuria per se temporarily reduces the capacity of the kidney to concentrate the urine. Thus, the maximum level of urine osmolarity achieved during fluid deprivation is often indistinguishable from that in patients with partial pituitary or partial nephrogenic DI.

Differential Diagnosis

When symptoms of urinary frequency, enuresis, nocturia, and/or persistent thirst are present in the absence of glucosuria, the possibility of DI should be evaluated by collecting a 24-h urine on ad libitum fluid intake. If the osmolarity is <300 mosmol/L and the volume >50 mL/kg per day, the patient has DI and should be evaluated further to determine the type and select appropriate therapy. If the volume and osmolarity are not concordant, the possibility of inaccurate collection can be evaluated by determining if total urinary creatinine is normal for the size of the patient (20–30 mg/kg/day).

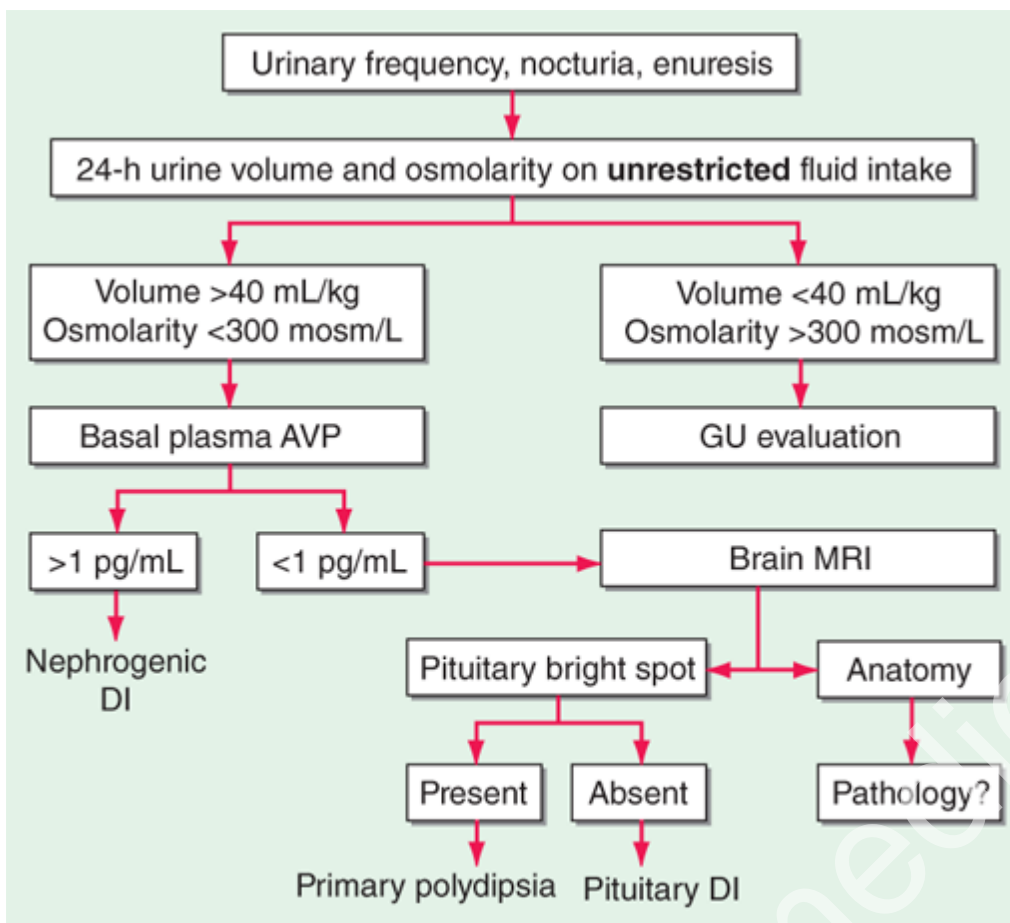
The type of DI can sometimes be inferred from the clinical setting or medical history. Often, however, such information is lacking, ambiguous, or misleading, and other approaches to differential diagnosis are needed. If basal plasma osmolarity and sodium are within normal limits, the traditional approach is to determine the effect of fluid deprivation and injection of antidiuretic hormone on urine osmolarity. This approach suffices for differential diagnosis *if* fluid deprivation raises plasma osmolarity and sodium above the normal range *without* inducing concentration of the urine. In that event, primary polydipsia and partial defects in AVP secretion and action are excluded, and the effect on urine osmolarity of injecting 2 μ g of the AVP analogue, desmopressin indicates whether the patient has severe pituitary DI or severe nephrogenic DI. However, this approach is of little or no diagnostic value if fluid deprivation results in concentration of the urine because the increases in urine osmolarity achieved both before and after the injection of desmopressin are similar in patients with *partial* pituitary DI, *partial* nephrogenic DI, and primary polydipsia. These disorders can be differentiated by measuring plasma AVP during fluid deprivation and relating it to the concurrent level of plasma and urine osmolarity (Fig. 374-3). However, this approach does not always differentiate clearly between partial pituitary DI and primary polydipsia unless the measurement is made when plasma osmolarity and sodium are at or above the normal range. This level is difficult to achieve by fluid deprivation alone once urinary concentration occurs. Therefore, it is usually necessary to give a short infusion of 3% saline condition (0.1 mL/kg body weight per minute for 60–90 min) and repeat the measurement of plasma AVP. This approach is highly reliable for differential diagnosis but it is often stressful for the patient and requires special facilities and staff to perform safely and accurately.

A simpler, and less stressful, but equally reliable way to differentiate between pituitary DI, nephrogenic DI, and primary polydipsia is to start by measuring basal plasma AVP and urine osmolarity under conditions of unrestricted fluid intake (Fig. 374-4). If AVP is normal or elevated (>1 pg/mL) and the concurrent urine osmolarity is low (<300 mosmol/L), the patient has nephrogenic DI and the only additional evaluation required is to determine the cause. If, however, basal plasma AVP is low or undetectable (<1 pg/mL), nephrogenic DI is very unlikely and MRI of the brain can be performed to differentiate pituitary DI from primary polydipsia. In most healthy adults and children, the posterior pituitary emits a hyperintense signal visible in T1-weighted

midsagittal images. This “bright spot” is almost always present in patients with primary polydipsia but is always absent or abnormally small in patients with pituitary DI, even if their AVP deficiency is partial. The MRI is also useful in searching for pathology responsible for pituitary DI or the dipsogenic form of primary polydipsia (Fig. 374-2). The principal caveat is that MRI is not reliable for differential diagnosis of DI in patients with empty sella because they typically lack a bright spot even when their AVP secretion and action are normal. MRI also cannot be used to differentiate pituitary from nephrogenic DI because many patients with nephrogenic DI also lack a posterior pituitary bright spot, probably because they have an abnormally high rate of AVP secretion and turnover.

FIGURE 374-4

Simplified approach to the differential diagnosis of diabetes insipidus. When symptoms suggest diabetes insipidus (DI), the syndrome should be differentiated from a genitourinary (GU) abnormality by measuring the 24-h urine volume and osmolality on unrestricted fluid intake. If DI is confirmed, basal plasma arginine vasopressin (AVP) should be measured on unrestricted fluid intake. If AVP is normal or elevated (>1 pg/mL), the patient probably has nephrogenic DI. However, if plasma AVP is low or undetectable, the patient has either pituitary DI or primary polydipsia. In that case, magnetic resonance imaging (MRI) of the brain can be performed to differentiate between these two conditions by determining whether or not the normal posterior pituitary bright spot is visible on T1-weighted midsagittal images. In addition, the MRI anatomy of the pituitary hypothalamic area can be examined to look for evidence of pathology that sometimes causes pituitary DI or the dipsogenic form of primary polydipsia. MRI is not reliable for differential diagnosis unless nephrogenic DI has been excluded because the bright spot is also absent, small, or faint in this condition.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

If MRI and/or AVP assays with the requisite sensitivity and specificity are unavailable and a fluid deprivation test is impractical or undesirable, a third way to differentiate between pituitary DI, nephrogenic DI, and primary polydipsia is a trial of desmopressin therapy. Such a trial should be conducted with very close monitoring of serum sodium as well as urine output, preferably in hospital, because desmopressin will produce hyponatremia in 8–24 h if the patient has primary polydipsia.

TREATMENT

TREATMENT

Diabetes Insipidus

The signs and symptoms of uncomplicated pituitary DI can be eliminated by treatment with desmopressin (DDAVP), a synthetic analogue of AVP (Fig. 374-1). DDAVP acts selectively at V2 receptors to increase urine concentration and decrease urine flow in a dose-dependent manner. It is also more resistant to degradation than is AVP and has a three- to fourfold longer duration of action. DDAVP can be given by IV or SC injection, nasal inhalation, or orally by means of a tablet or melt. The doses required to control pituitary DI vary widely, depending on the patient and the route of administration. However, among adults, they usually range from 1–2 µg qd or bid by injection, 10–20 µg bid or tid by nasal spray, or 100–400 µg bid or tid orally. The onset of antidiuresis is rapid, ranging from as little as 15 min after injection to 60 min after oral administration. When given in a dose that normalizes 24-h urinary osmolarity (400–800 mosmol/L) and volume (15–30 mL/kg body

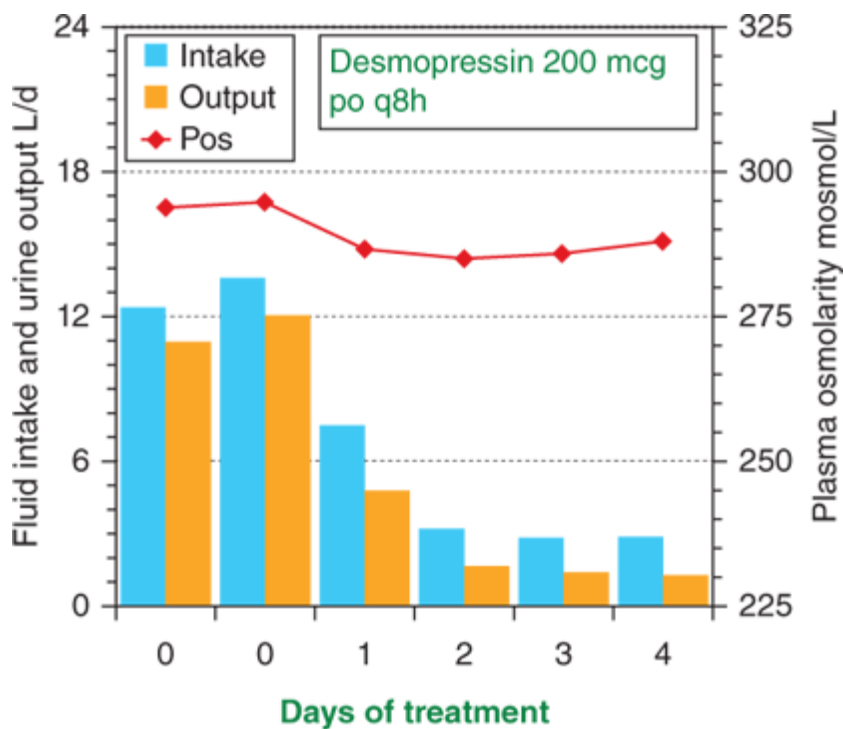
weight), DDAVP produces a slight (1–3%) increase in total body water and a decrease in plasma osmolarity/sodium that rapidly eliminates thirst and polydipsia (**Fig. 374-5**). Consequently, water balance is maintained within the normal range. Hyponatremia rarely develops unless urine volume is reduced too far (to <10 mL/kg per day) or fluid intake is excessive due to an associated abnormality in thirst or cognition. Fortunately, thirst abnormalities are rare, and if the patient is taught to drink only when truly thirsty, DDAVP can be given safely in doses sufficient to normalize urine output without the need for allowing intermittent escape to prevent water intoxication.

Primary polydipsia cannot be treated safely with DDAVP or any other antidiuretic drug because eliminating the polyuria does not eliminate the urge to drink. Therefore, it invariably produces hyponatremia and/or other signs of water intoxication, usually within 8–24 h if urine output is normalized completely. There is no consistently effective way to correct dipsogenic or psychogenic polydipsia, but the iatrogenic form may respond to patient education. To minimize the risk of water intoxication, all patients should be warned about the use of other drugs such as thiazide diuretics or carbamazepine (Tegretol) that can impair urinary free-water excretion directly or indirectly.

The polyuria and polydipsia of nephrogenic DI are not affected by treatment with standard doses of DDAVP. If resistance is partial, it may be overcome by tenfold higher doses, but this treatment is too expensive and inconvenient for long-term use. However, treatment with conventional doses of a thiazide diuretic and/or amiloride in conjunction with a low-sodium diet and a prostaglandin synthesis inhibitor (e.g., **indomethacin**) usually reduces the polyuria and polydipsia by 30–70% and may eliminate them completely in some patients. Side effects such as hypokalemia and gastric irritation can be minimized by the use of amiloride or potassium supplements and by taking medications with meals.

FIGURE 374-5

Effect of desmopressin therapy on fluid intake (*blue bars*), urine output (*orange bars*), and plasma osmolarity (*red line*) in a patient with uncomplicated pituitary diabetes insipidus. Note that treatment rapidly reduces fluid intake and urine output to normal, with only a slight increase in body water as evidenced by the slight decrease in plasma osmolarity.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

HYPODIPSIC HYPERNATREMIA

An increase in plasma osmolality/sodium above the normal range (hypertonic hypernatremia) can be caused by either a decrease in total body water or an increase in total body sodium. The former results from a failure to drink enough to replace normal or increased urinary and insensible water loss. The deficient intake can be due either to water deprivation or a lack of thirst (hypodipsia). The most common cause of an increase in total body sodium is primary hyperaldosteronism ([Chap. 379](#)). Rarely, it can also result from ingestion of hypertonic saline in the form of sea water or incorrectly prepared infant formula. However, even in these forms of hypernatremia, inadequate intake of water also contributes. This chapter focuses on hypodipsic hypernatremia, the form of hypernatremia due to a primary defect in the thirst mechanism.

Clinical Characteristics

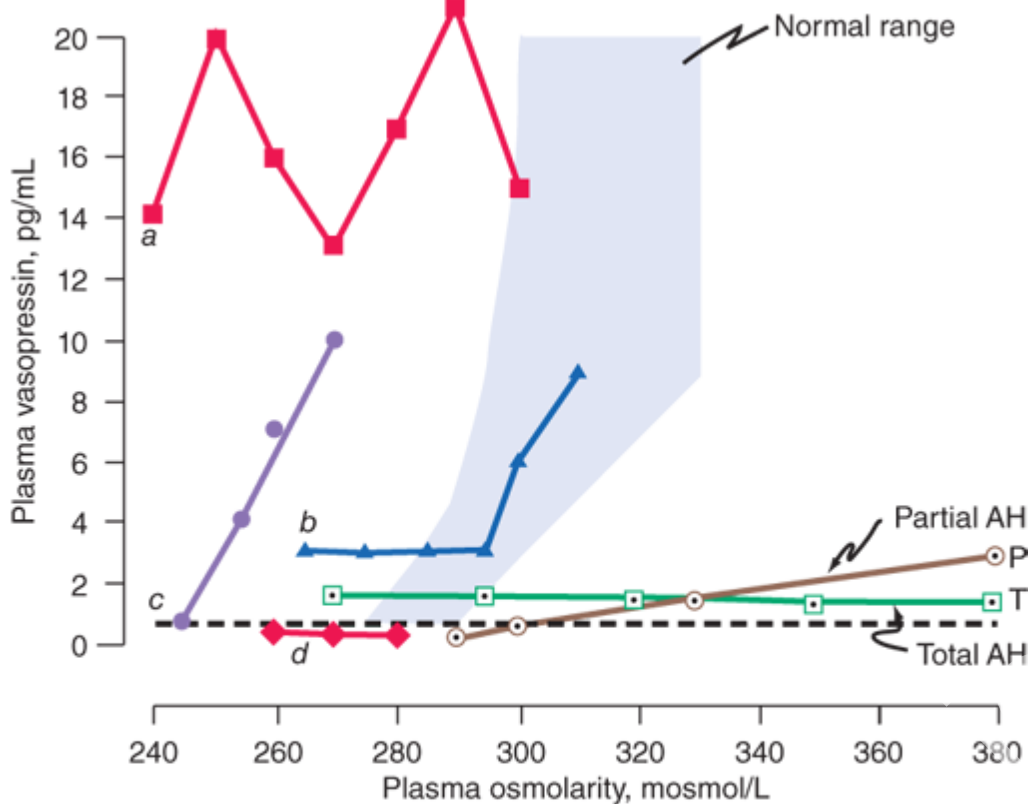
Hypodipsic hypernatremia is a syndrome characterized by chronic or recurrent hypertonic dehydration. The hypernatremia varies widely in severity and usually is associated with signs of hypovolemia such as tachycardia, postural hypotension, azotemia, hyperuricemia, and hypokalemia due to secondary hyperaldosteronism. Muscle weakness, pain, rhabdomyolysis, hyperglycemia, hyperlipidemia, and acute renal failure may also occur. Obtundation or coma may be present but are often absent. Despite inappropriately low levels of plasma AVP, DI usually is not evident at presentation but may develop during rehydration as blood volume, blood pressure, and plasma osmolality/sodium return toward normal, further reducing plasma AVP.

Etiology

Hypodipsia is usually due to hypogenesis or destruction of the osmoreceptors in the anterior hypothalamus that regulate thirst. These defects can result from various congenital malformations of midline brain structures or may be acquired due to diseases such as occlusions of the anterior communicating artery, primary, or metastatic tumors in the hypothalamus, head trauma, surgery, granulomatous diseases such as sarcoidosis and histiocytosis, AIDS, and cytomegalovirus encephalitis. Because of their proximity, the osmoreceptors that regulate AVP secretion also are usually impaired. Thus, AVP secretion responds poorly or not at all to hyperosmotic stimulation (Fig. 374-6) but, in most cases, increases normally to nonosmotic stimuli such as nausea or large reductions in blood volume or blood pressure, indicating that the neurohypophysis is intact.

FIGURE 374-6

Heterogeneity of osmoregulatory dysfunction in adipsic hypernatremia (AH) and the syndrome of inappropriate antidiuresis (SIAD). Each line depicts schematically the relationship of plasma arginine vasopressin (AVP) to plasma osmolarity during water loading and/or infusion of 3% saline in a patient with either AH (*open symbols*) or SIAD (*closed symbols*). The shaded area indicates the normal range of the relationship. The horizontal broken line indicates the plasma AVP level below which the hormone is undetectable and urinary concentration usually does not occur. Lines P and T represent patients with a selective deficiency in the osmoregulation of thirst and AVP that is either partial (⊙) or total (◻). In the latter, plasma AVP does not change in response to increases or decreases in plasma osmolarity but remains within a range sufficient to concentrate the urine even if overhydration produces hypotonic hyponatremia. In contrast, if the osmoregulatory deficiency is partial (⊙), rehydration of the patient suppresses plasma AVP to levels that result in urinary dilution and polyuria before plasma osmolarity and sodium are reduced to normal. Lines *a–d* represent different defects in the osmoregulation of plasma AVP observed in patients with SIADH or SIAD. In *a* (■), plasma AVP is markedly elevated and fluctuates widely without relation to changes in plasma osmolarity, indicating complete loss of osmoregulation. In *b* (▲), plasma AVP remains fixed at a slightly elevated level until plasma osmolarity reaches the normal range, at which point it begins to rise appropriately, indicating a selective defect in the inhibitory component of the osmoregulatory mechanism. In *c* (●), plasma AVP rises in close correlation with plasma osmolarity before the latter reaches the normal range, indicating downward resetting of the osmostat. In *d* (◆), plasma AVP appears to be osmoregulated normally, suggesting that the inappropriate antidiuresis is caused by some other abnormality.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Pathophysiology

Hypodipsia results in a failure to drink enough water to replenish obligatory renal and extrarenal losses. Consequently, plasma osmolarity and sodium rise often to extremely high levels before the disorder is recognized. In most cases, urinary loss of water contributes little, if any, to the dehydration because AVP continues to be secreted in the small amounts necessary to concentrate the urine. In some patients this appears to be due to hypovolemic stimulation and/or incomplete destruction of AVP osmoreceptors because plasma AVP declines and DI develops during rehydration (Fig. 374-6). In others, however, plasma AVP does not decline during rehydration even if they are overhydrated. Consequently, they develop a hyponatremic syndrome indistinguishable from inappropriate antidiuresis. This suggests that the AVP osmoreceptors normally provide inhibitory and stimulatory input to the neurohypophysis and the patients can no longer osmotically stimulate or suppress tonic secretion of the hormone because both inputs have been totally eliminated by the same pathology that destroyed the osmoregulation of thirst. In a few patients, the neurohypophysis is also destroyed, resulting in a combination of chronic pituitary DI and hypodipsia that is particularly difficult to manage.

Differential Diagnosis

Hypodipsic hypernatremia usually can be distinguished from other causes of inadequate fluid intake (e.g., coma, paralysis, restraints, absence of fresh water) by the clinical history and setting. Previous episodes and/or denial of thirst and failure to drink spontaneously when the patient is conscious, unrestrained, and

hypernatremic are virtually diagnostic. The hypernatremia caused by excessive retention or intake of sodium can be distinguished by the presence of thirst as well as the physical and laboratory signs of hypervolemia rather than hypovolemia.

TREATMENT

TREATMENT

Hypodipsic Hypernatremia

Hypodipsic hypernatremia should be treated by administering water orally if the patient is alert and cooperative or by infusing hypotonic fluids (0.45% saline or 5% dextrose and water) if the patient is not. The amount of free water in liters required to correct the deficit (ΔFW) can be estimated from body weight in kg (BW) and the serum sodium concentration in mmol/L (S_{Na}) by the formula $\Delta FW = 0.5BW \times ([S_{Na} - 140]/140)$. If serum glucose (S_{Glu}) is elevated, the measured S_{Na} should be corrected (S_{Na}^*) by the formula $S_{Na}^* = S_{Na} + ([S_{Glu} - 90]/36)$. This amount plus an allowance for continuing insensible and urinary losses should be given over a 24- to 48-h period. Close monitoring of serum sodium as well as fluid intake and urinary output is essential because, depending on the extent of osmoreceptor deficiency, some patients will develop AVP-deficient DI, requiring DDAVP therapy to complete rehydration; others will develop hyponatremia and a syndrome of inappropriate antidiuresis (SIAD)-like picture if overhydrated. If hyperglycemia and/or hypokalemia are present, insulin and/or potassium supplements should be given with the expectation that both can be discontinued soon after rehydration is complete. Plasma urea/creatinine should be monitored closely for signs of acute renal failure caused by rhabdomyolysis, hypovolemia, and hypotension.

Once the patient has been rehydrated, an MRI of the brain and tests of anterior pituitary function should be performed to look for the cause and collateral defects in other hypothalamic functions. A long-term management plan to prevent or minimize recurrence of the fluid and electrolyte imbalance also should be developed. This should include a practical method to regulate fluid intake in accordance with variations in water balance as indicated by changes in body weight or serum sodium determined by home monitoring analyzers. Prescribing a constant fluid intake is ineffective and potentially dangerous because it does not take into account the large, uncontrolled variations in insensible loss that inevitably result from changes in ambient temperature and physical activity.

HYPONATREMIA DUE TO INAPPROPRIATE ANTIDIURESIS

A decrease in plasma osmolarity/sodium below the normal range (hypotonic hyponatremia) can be due to any of three different types of salt and water imbalance: (1) an increase in total body water that exceeds the increase in total body sodium (hypervolemic hyponatremia); (2) a decrease in body sodium greater than the decrease in body water (hypovolemic hyponatremia); or (3) an increase in body water with little or no change in body sodium (euvolemic hyponatremia) (**Chap. 49**). All three forms are associated with a failure to fully dilute the urine and mount a water diuresis in the face of hypotonic hyponatremia. However, the disorders with which they are associated and the types of salt and water imbalance that result differ. The hypervolemic

form typically occurs in disorders like severe congestive heart failure or cirrhosis in which water is retained in excessive of sodium. The hypovolemic form typically occurs in disorders such as severe diarrhea, diuretic abuse, or mineralocorticoid deficiency in which sodium is lost in excess of water. Euvolemic hyponatremia, however, is due mainly to expansion of total body water caused by excessive intake in the face of a failure to dilute the urine in response to excessive water intake. The impaired dilution is usually caused by a defect in the osmotic suppression of AVP that can have either of two causes. One is a nonhemodynamic stimulus such as nausea or a cortisol deficiency, which can be corrected quickly by treatment with antiemetics or cortisol. The other is a primary defect in osmoregulation caused by another disorder such as malignancy, stroke, or pneumonia that cannot be easily or quickly corrected. The latter is commonly known as the syndrome of inappropriate antidiuretic hormone (SIADH). Much less often, euvolemic hyponatremia can also result from AVP-independent activation of renal V_2 receptors, a variant known as nephrogenic inappropriate antidiuresis or NSIAD. Both of the latter will be discussed in this chapter.

Clinical Characteristics

Antidiuresis of any cause decreases the volume and increases the concentration of urine. If not accompanied by a commensurate reduction in fluid intake or an increase in insensible loss, the reduction in urine output results in excess water retention which expands and dilutes body fluids. If the hyponatremia develops gradually or has been present for more than a few days, it may be largely asymptomatic. However, if it develops acutely, it is usually accompanied by symptoms and signs of water intoxication that may include mild headache, confusion, anorexia, nausea, vomiting, coma, and convulsions. Severe acute hyponatremia may be lethal. Other clinical signs and symptoms vary greatly, depending on the type of hyponatremia. The hypervolemic form is characterized by generalized edema and other signs of marked volume expansion. The opposite is evident in the hypovolemic form. However, overt signs of volume expansion or contraction are absent in SIADH, SIAD, NSIAD, and other forms of euvolemic hyponatremia.

Etiology

In SIADH, the inappropriate secretion of AVP can have many different causes. They include ectopic production of AVP by lung cancer or other neoplasms; eutopic release induced by various diseases or drugs; and exogenous administration of AVP, DDAVP, or large doses of oxytocin ([Table 374-2](#)). The ectopic forms result from abnormal expression of the *AVP-NPII* gene by primary or metastatic malignancies. The eutopic forms occur most often in patients with acute infections or strokes but have also been associated with many other neurologic diseases and injuries. The mechanisms by which these diseases interfere with osmotic suppression of AVP are not known. The defect in osmoregulation can take any of four distinct forms ([Fig. 374-6](#)). In one of the most common (reset osmostat), AVP secretion remains fully responsive to changes in plasma osmolality/sodium, but the threshold, or set point, of the osmoregulatory system is abnormally low. These patients differ from those with the other types of SIADH in that they are able to maximally suppress plasma AVP and dilute their urine if their fluid intake is high enough to reduce their plasma osmolality and/or sodium to the lower set point. In most patients, SIADH is self-limited and remits spontaneously within 2–3 weeks, but about 10% of cases are chronic. Another, smaller subgroup (~10% of the total) has inappropriate

antidiuresis without a demonstrable defect in the osmoregulation of plasma AVP (Fig. 374-6). In some of them, all young boys, the inappropriate antidiuresis has been traced to a constitutively activating mutation of the V_2 receptor gene. This unusual variant may be referred to as familial nephrogenic SIAD (NSIAD) to distinguish it from other possible causes of the syndrome. The inappropriate antidiuresis in these patients appears to be permanent, although the hyponatremia is variable owing presumably to individual differences in fluid intake.

TABLE 374-2

Causes of Syndrome of Inappropriate Antidiuretic Hormone (SIADH)

Neoplasms	Neurologic
Carcinomas	Guillain-Barré syndrome
Lung	Multiple sclerosis
Duodenum	Delirium tremens
Pancreas	Amyotrophic lateral sclerosis
Ovary	Hydrocephalus
Bladder, ureter	Psychosis
Other neoplasms	Peripheral neuropathy
Thymoma	Congenital malformations
Mesothelioma	Agenesis corpus callosum
Bronchial adenoma	Cleft lip/palate
Carcinoid	Other midline defects
Gangliocytoma	Metabolic
Ewing's sarcoma	Acute intermittent porphyria
Head trauma (closed and penetrating)	Pulmonary
Infections	Asthma
Pneumonia, bacterial or viral	Pneumothorax
Abscess, lung or brain	Positive-pressure respiration
Cavitation (aspergillosis)	Drugs
Tuberculosis, lung or brain	Vasopressin or desmopressin
Meningitis, bacterial or viral	Serotonin reuptake inhibitors
Encephalitis	Oxytocin, high dose
AIDS	Vincristine
Vascular	Carbamazepine
Cerebrovascular occlusions, hemorrhage	Nicotine
Cavernous sinus thrombosis	Phenothiazines
	Cyclophosphamide
	Tricyclic antidepressants
	Monoamine oxidase inhibitors

Pathophysiology

Impaired osmotic suppression of antidiuresis results in excessive retention of water and dilution of body fluids only if water intake exceeds insensible and urinary losses. The excess intake is sometimes due to an associated defect in the osmoregulation of thirst (dipsogenic) but can also be psychogenic or iatrogenic,

including excessive IV administration of hypotonic fluids. In SIADH and other forms of euvolemic hyponatremia, the decrease in plasma osmolarity/sodium and the increase in extracellular and intracellular volume are proportional to the amount of water retained. Thus, an increase in body water of 10% (~4 L in a 70-kg adult) reduces plasma osmolarity and sodium by ~10% (~28 mosmol/L or 14 meq/L). An increase in body water of this magnitude is rarely detectable on physical examination but will be reflected in a weight gain of about 4 kg. It also increases glomerular filtration and atrial natriuretic hormone and suppresses plasma renin activity, thereby increasing urinary sodium excretion. The resultant reduction in total body sodium decreases the expansion of extracellular volume but aggravates the hyponatremia and further expands intracellular volume. The latter increases brain swelling and intracranial pressure, which probably produces most of the symptoms of acute water intoxication. Within a few days, this swelling may be counteracted by inactivation or elimination of intracellular solutes, resulting in the remission of symptoms even though the hyponatremia persists.

In type I (hypervolemic) or type II (hypovolemic) hyponatremia, osmotic suppression of AVP secretion appears to be counteracted by a hemodynamic stimulus resulting from a large reduction in cardiac output and/or effective blood volume. The resultant antidiuresis is enhanced by decreased distal delivery of glomerular filtrate that results from increased reabsorption of sodium in proximal nephron. If the reduction in urine output is not associated with a commensurate reduction in water intake or an increase in insensible loss, body fluids are expanded and diluted, resulting in hyponatremia despite an increase in body sodium. Unlike SIADH and other forms of euvolemic hyponatremia, however, glomerular filtration is reduced and plasma renin activity and aldosterone are elevated. Thus, the rate of urinary sodium excretion is low (unless sodium reabsorption is impaired by a diuretic), and the hyponatremia is usually accompanied by edema, hypokalemia, azotemia, and hyperuricemia. In type II (hypovolemic) hyponatremia, sodium and water are also retained as an appropriate compensatory response to the severe depletion.

Differential Diagnosis

SIADH is a diagnosis of exclusion that usually can be made from the history, physical examination, and basic laboratory data. If hyperglycemia is present, its contribution to the reduction in plasma sodium can be estimated either by measuring plasma osmolarity for a more accurate estimate of the true “effective” tonicity of body fluids or by correcting the measured plasma sodium for the reduction caused by the hyperglycemia using the simplified formula

$$\text{corrected } P_{\text{Na}} = \text{measured } P_{\text{Na}} + (P_{\text{glu}} - 90)/36$$

where P_{Na} = plasma sodium in meq/L and P_{glu} = plasma glucose in mg/dL.

If the plasma osmolarity and/or corrected plasma sodium are below normal limits, hypotonic hyponatremia is present and further evaluation to determine the type should be undertaken in order to administer safe and effective treatment. This differentiation is usually possible by evaluating standard clinical indicators of the extracellular fluid volume ([Table 374-3](#)). If these findings are ambiguous or contradictory, measuring plasma

renin activity or the *rate* of urinary sodium excretion may be helpful *provided* that the hyponatremia is not in the recovery phase or is due to a primary defect in renal conservation of sodium, diuretic abuse, or hyporeninemic hypoaldosteronism. The latter may be suspected if serum potassium is elevated instead of low, as it usually is in types I and II hyponatremia. Measurements of plasma AVP are currently of no value in differentiating SIADH from the other types of hyponatremia since the plasma levels are elevated similarly in all. In patients who fulfill the clinical criteria for type III (euvoletic) hyponatremia, morning plasma cortisol should also be measured to exclude secondary adrenal insufficiency. If it is normal and there is no history of nausea/vomiting, the diagnosis of SIADH is confirmed, and a careful search for occult lung cancer or other common causes of the syndrome ([Table 374-2](#)) should be undertaken.

TABLE 374-3

Differential Diagnosis of Hyponatremia Based on Clinical Assessment of Extracellular Fluid Volume (ECFV)

CLINICAL FINDINGS	TYPE I, HYPERVOLEMIC	TYPE II, HYPOVOLEMIC	TYPE III, EUVOLEMIC	SIADH AND SIAD EUVOLEMIC
History				
CHF, cirrhosis, or nephrosis	Yes	No	No	No
Salt and water loss	No	Yes	No	No
ACTH–cortisol deficiency and/or nausea and vomiting	No	No	Yes	No
Physical examination				
Generalized edema, ascites	Yes	No	No	No
Postural hypotension	Maybe	Maybe	Maybe ^a	No
Laboratory				
BUN, creatinine	High-normal	High-normal	Low-	Low-normal
Uric acid	High-normal	High-normal	normal	Low-normal
Serum potassium	Low-normal	Low-normal ^b	Low-	Normal
Serum urate	High	High	normal	Low
Serum albumin	Low-normal	High-normal	Normal ^c	Normal
Serum cortisol	Normal-high	Normal-high ^d	Low	Normal
Plasma renin activity	High	High	Normal	Low
Urinary sodium (meq per unit of time) ^g	Low	Low ^h	Low ^e Low ^f High ⁱ	High ⁱ

^aPostural hypotension may occur in secondary (ACTH-dependent) adrenal insufficiency even though extracellular fluid volume and aldosterone are usually normal. ^bSerum potassium may be high if hypovolemia is due to aldosterone deficiency. ^cSerum potassium may be low if vomiting causes alkalosis. ^dSerum cortisol is low if hypovolemia is due to primary adrenal insufficiency (Addison's disease). ^eSerum cortisol will be normal or high if the cause is nausea and vomiting rather than secondary (ACTH-dependent) adrenal insufficiency. ^fPlasma renin activity may be high if the cause is secondary (ACTH) adrenal insufficiency. ^gUrinary sodium should be expressed as the *rate of excretion* rather than the concentration. In a hyponatremic adult, an excretion rate >25 meq/d (or 25 µeq/mg of creatinine) could be considered high. ^hThe rate of urinary sodium excretion may be high if the hypovolemia is due to diuretic abuse,

primary adrenal insufficiency, or other causes of renal sodium wasting.ⁱ The rate of urinary sodium excretion may be low if intake is curtailed by symptoms or treatment.

Abbreviations: ACTH, adrenocorticotrophic hormone; BUN, blood urea nitrogen; CHF, congestive heart failure; SIAD, syndrome of inappropriate antidiuresis.

SIAD due to an activating mutation of the V₂ receptor gene should be suspected if the hyponatremia occurs in a child or several members of the family or is refractory to treatment with a vaptan (see below). In that case, plasma AVP should be measured to confirm that it is appropriately suppressed while the hyponatremia and antidiuresis are present, and the V₂ receptor gene should be sequenced, if possible.

TREATMENT

TREATMENT

Hyponatremia

The management of hyponatremia differs depending on the type and the severity and duration of symptoms. In acute symptomatic SIADH, the aim should be to raise plasma osmolality and/or plasma sodium at a rate ~1% an hour until they reach levels of ~270 mosmol/L or 130 meq/L, respectively. This can be accomplished in either of two ways. One is to infuse hypertonic (3%) saline at a rate of about 0.05 mL/kg body weight per min. This treatment often produces a solute diuresis that serves to remove some of the excess water. The other treatment for acute, symptomatic SIADH is to reduce body water by giving an AVP receptor-2 antagonist (vaptan) to block the antidiuretic effect of AVP and increase urine output (**Fig. 374-7**). One of the vaptans, a combined V₂/V_{1a} antagonist (Conivaptan), has been approved for short-term, in-hospital IV treatment of SIADH. It should be given as a loading dose of 20 mg IV over 30 min followed by a continuous infusion of 20 mg over 24 h. Another vaptan (Tolvaptan) can be given orally starting at a dose of 15 mg po and increasing to 30 mg or 60 mg at 24 h intervals depending on the effect. With either approach, fluid intake should be restricted to less than urine output. Because the aquaretic effect of the vaptans varies in magnitude from patient to patient, the rate of rise in serum sodium also varies if fluid intake is fixed at a constant rate. This variability in effect can be reduced or eliminated by continuously monitoring the rate of urine output and adjusting the rate of IV or oral fluid intake so as to reduce body water at a constant rate. Regulating fluid intake so that it under replaces urine output by 5mL/kg body weight/h will raise serum sodium at a rate of about 1% an hour. In any event, serum sodium should be checked every 2–4 h to ensure it is not raised faster than 1mEq/L per hour or above the lower limit of normal. Doing so may result in central pontine myelinolysis, an acute, potentially fatal neurologic syndrome characterized by quadriparesis, ataxia, and abnormal extraocular movements.

In chronic and/or minimally symptomatic SIADH, the hyponatremia can and should be corrected more gradually. This can be achieved by restricting total fluid intake to less than the sum of urinary and insensible losses. Because the water derived from food (300–700 mL/d) usually approximates basal insensible losses in adults, the aim should be to reduce total discretionary intake (all liquids) to ~500 mL less than urinary

output. Adherence to this regimen is often problematic and, even if achieved, usually reduces body water and increases serum sodium by only about 1–2% per day. Therefore, it is often necessary to add a treatment that increases urinary water excretion. The oral AVP2 antagonist, tolvaptan, is best suited for this purpose. The best approach for treatment of chronic SIADH is the administration of an oral vaptan, tolvaptan, a selective V_2 antagonist that also increases urinary water excretion by blocking the antidiuretic effect of AVP. Some restriction of fluid intake may also be necessary to achieve satisfactory control of the hyponatremia. It is approved for treatment of nonemergent SIADH with initial in-hospital dosing. Other approaches include [demeclocycline](#), 150–300 mg PO tid or qid, which induces a reversible form of nephrogenic DI in 1–2 weeks, or fludrocortisone, 0.05–0.2 mg PO bid. The effect of the [demeclocycline](#) manifests in 7–14 days and is due to induction of a reversible form of nephrogenic DI. Fludrocortisone, 0.05–0.2 mg po bid, also raises serum sodium gradually over 1–2 weeks. Its mechanism of action is unclear but probably involves increased retention of sodium. It also increases urinary potassium excretion, which may require replacement through dietary adjustments or supplements and may induce hypertension, occasionally necessitating discontinuation of the treatment.

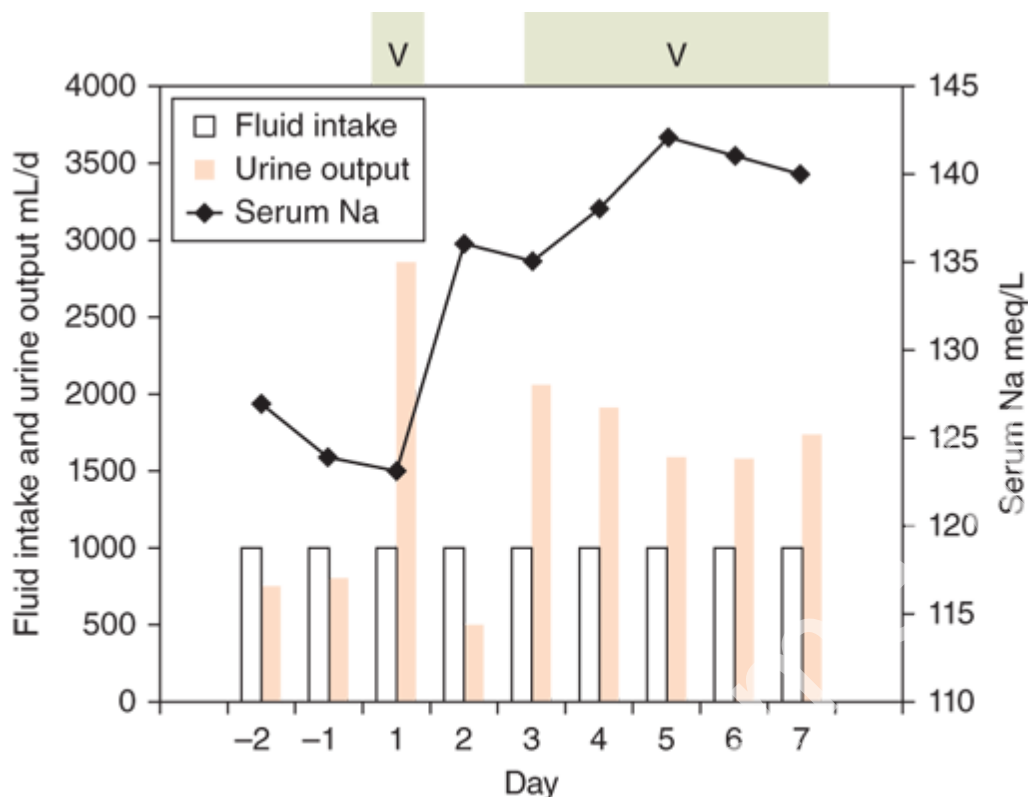
In the type of euvolemic hyponatremia caused by protracted nausea and vomiting or isolated glucocorticoid deficiency (type III), all abnormalities can be corrected quickly and completely by giving an antiemetic or stress doses of [hydrocortisone](#) (for glucocorticoid deficiency). As with other treatments, care must be taken to ensure that serum sodium does not rise too quickly or too far.

In SIAD due to an activating mutation of the V_2 receptor, the V_2 antagonists may not block the antidiuresis or raise plasma osmolarity/sodium. In that condition, use of an osmotic diuretic such as [urea](#) is reported to be effective in preventing or correcting hyponatremia. However, some vaptans may be effective in patients with a different type of activating mutation so the response to this therapy may be neither predictable nor diagnostic.

In hypervolemic hyponatremia, fluid restriction is also appropriate and somewhat effective if it can be maintained. The infusion of hypertonic saline is contraindicated because it further increases total body sodium and edema and may precipitate cardiovascular decompensation. However, as in SIADH, the V_2 receptor antagonists are also safe and effective in the treatment of hypervolemic hyponatremia caused by congestive heart failure. Tolvaptan is approved by the Food and Drug Administration for this indication with the caveat that treatment should be initiated or reinitiated in hospital. Its use should also be limited to 30 days at a time because of reports that longer periods may be associated with abnormal liver chemistries.

In hypovolemic hyponatremia, the imbalance can be corrected easily and quickly by stopping the loss of sodium and water and/or replacing the deficits by mouth or IV infusion of normal or hypertonic saline. As with the treatment of other forms of hyponatremia, care must be taken to ensure that plasma sodium does not increase too rapidly or too far. Fluid restriction and administration of AVP antagonists are contraindicated in type II hyponatremia because they would only aggravate the underlying volume depletion and could result in hemodynamic collapse.

The effect of vaptan therapy on water balance in a patient with chronic syndrome of inappropriate antidiuretic hormone (SIADH). The periods of vaptan (V) therapy are indicated by the green shaded boxes at the top. Urine output is indicated by orange bars. Fluid intake is shown by the open bars. Intake was restricted to 1 L/d throughout. Serum sodium is indicated by the black line. Note that sodium increased progressively when vaptan increased urine output to levels that clearly exceeded fluid intake.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

GLOBAL PERSPECTIVES



The incidence, clinical characteristics, etiology, pathophysiology, differential diagnosis, and treatments of fluid and electrolyte disorders in tropical and nonindustrialized countries differ in some respects from those in the United States and other industrialized parts of the world. Hyponatremia, for example, appears to be more common and is more likely to be due to infectious diseases such as cholera, shigellosis, and other diarrheal disorders. In these circumstances, hyponatremia is probably due to gastrointestinal losses of salt and water (hypovolemia type II), but other abnormalities, including undefined infectious toxins, also may contribute. The causes of DI are similar worldwide except that malaria and venoms from snake or insect bites are much more common in some tropical climates.

FURTHER READING

Robertson GL: Vaptans for the treatment of hyponatremia. *Nature Rev Endocrin* 7:151, 2011.

[CrossRef](#)

Robertson GL: Diabetes insipidus: Differential diagnosis and management. Best Prac Res Clin Endocrin Metabol 30:205, 2016.

[CrossRef](#)

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 375: Thyroid Gland Physiology and Testing

J. Larry Jameson; Susan J. Mandel; Anthony P. Weetman

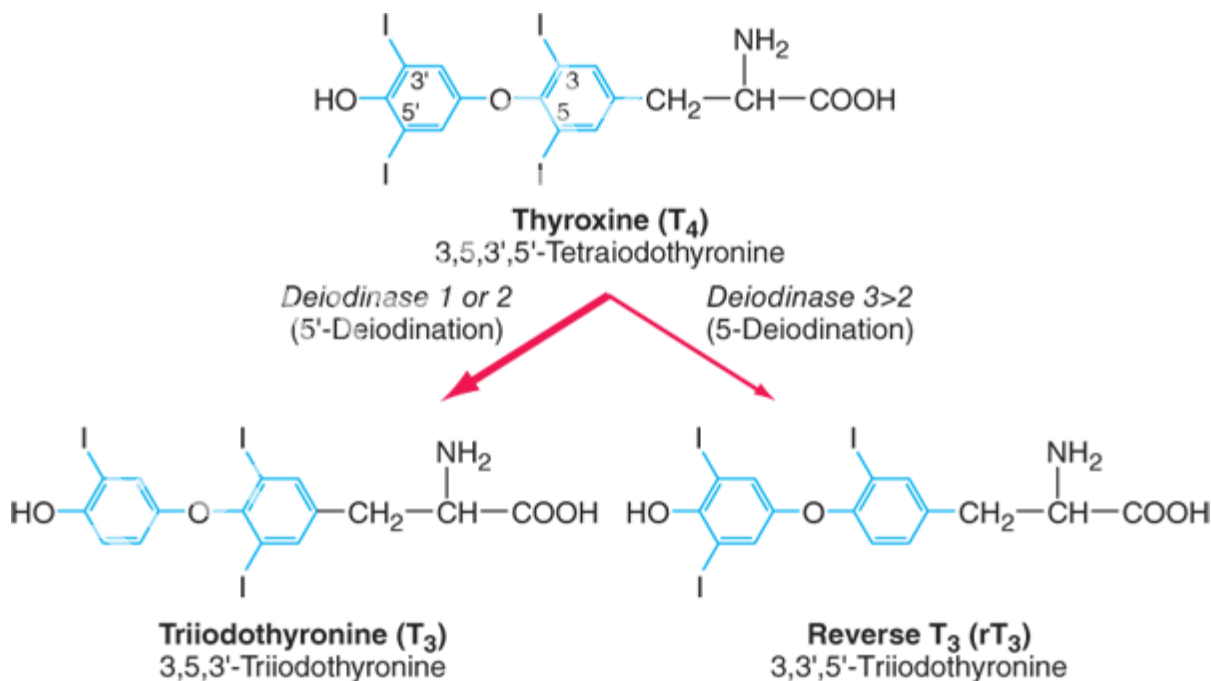
INTRODUCTION

The thyroid gland produces two related hormones, thyroxine (T_4) and triiodothyronine (T_3) (**Fig. 375-1**).

Acting through thyroid hormone receptors α and β , these hormones play a critical role in cell differentiation and organogenesis during development and help maintain thermogenic and metabolic homeostasis in the adult. Autoimmune disorders of the thyroid gland can stimulate overproduction of thyroid hormones (*thyrotoxicosis*) or cause glandular destruction and hormone deficiency (*hypothyroidism*). Benign nodules and various forms of thyroid cancer are relatively common and amenable to detection by physical examination or various imaging techniques.

FIGURE 375-1

Structures of thyroid hormones. Thyroxine (T_4) contains four iodine atoms. Deiodination leads to production of the potent hormone triiodothyronine (T_3) or the inactive hormone reverse T_3 .



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

ANATOMY AND DEVELOPMENT

The thyroid (Greek *thyreos*, shield, plus *eidos*, form) consists of two lobes connected by an isthmus. It is located anterior to the trachea between the cricoid cartilage and the suprasternal notch. The normal thyroid is 12–20 g in size, highly vascular, and soft in consistency. Four parathyroid glands, which produce parathyroid hormone (**Chap. 403**), are located posterior to each pole of the thyroid. The recurrent laryngeal nerves traverse the lateral borders of the thyroid gland and must be identified during thyroid surgery to avoid injury and vocal cord paralysis.

The thyroid gland develops from the floor of the primitive pharynx during the third week of gestation. The developing gland migrates along the thyroglossal duct to reach its final location in the neck. This feature accounts for the rare ectopic location of thyroid tissue at the base of the tongue (lingual thyroid) as well as the occurrence of thyroglossal duct cysts along this developmental tract. Thyroid hormone synthesis normally begins at about 11 weeks' gestation.

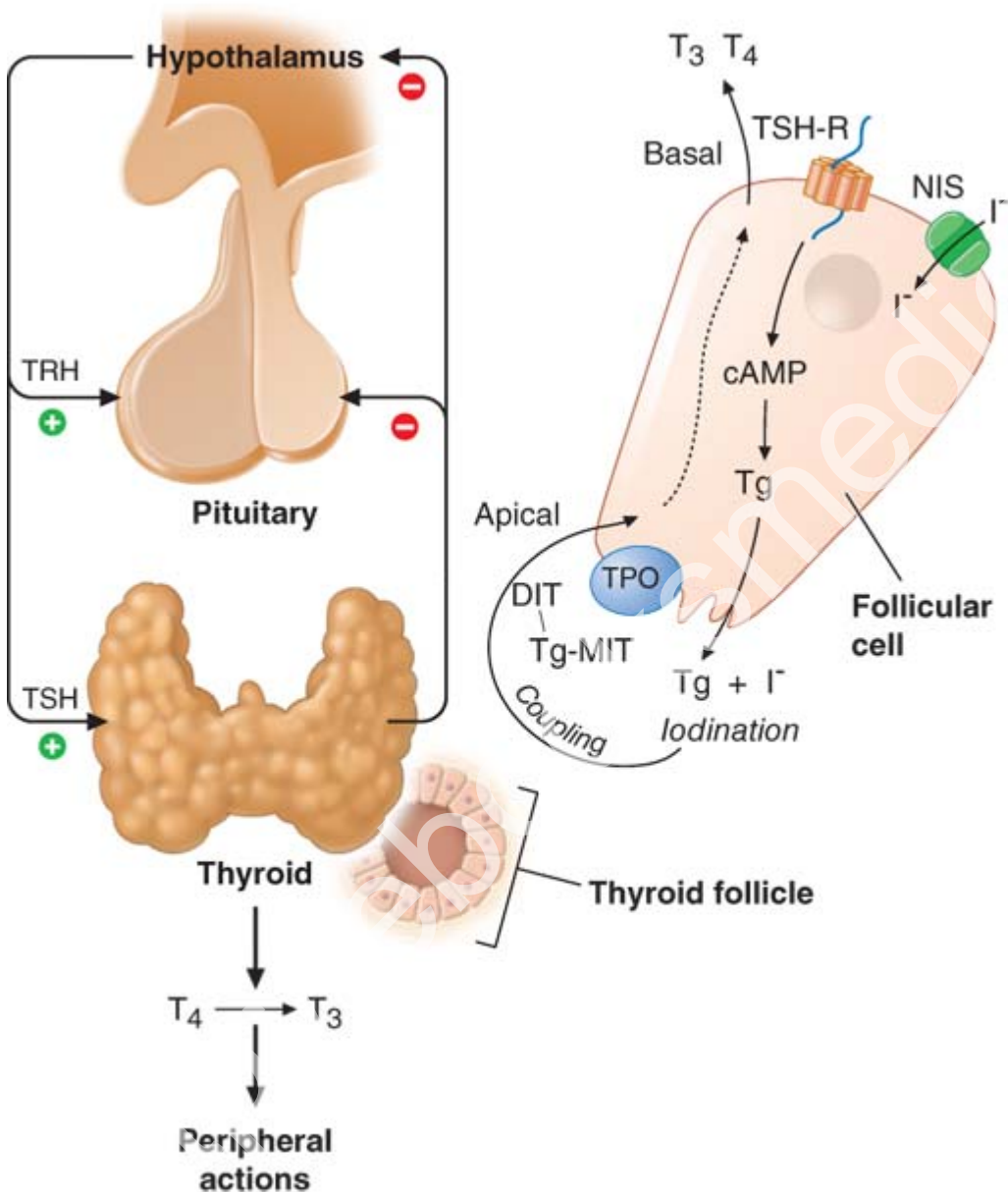
Neural crest derivatives from the ultimobranchial body give rise to thyroid medullary C cells that produce calcitonin, a calcium-lowering hormone. The C cells are interspersed throughout the thyroid gland, although their density is greatest in the juncture of the upper one-third and lower two-thirds of the gland. Calcitonin plays a minimal role in calcium homeostasis in humans but the C-cells are important because of their involvement in medullary thyroid cancer.

Thyroid gland development is orchestrated by the coordinated expression of several developmental transcription factors. Thyroid transcription factor (TTF)-1, TTF-2, NKX2-1, and paired homeobox-8 (PAX-8) are expressed selectively, but not exclusively, in the thyroid gland. In combination, they dictate thyroid cell development and the induction of thyroid-specific genes such as thyroglobulin (Tg), thyroid peroxidase (TPO), the sodium iodide symporter (Na^+/I^- , NIS), and the thyroid-stimulating hormone receptor (TSH-R). Mutations in these developmental transcription factors or their downstream target genes are rare causes of thyroid agenesis or dysmorphogenesis, although the causes of most forms of congenital hypothyroidism remain unknown (**see Chap. 376, Table 376-1**). Because congenital hypothyroidism occurs in ~1 in 4000 newborns, neonatal screening is now performed in most industrialized countries. Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides significant hormone support to a fetus with congenital hypothyroidism. Early thyroid hormone replacement in newborns with congenital hypothyroidism prevents potentially severe developmental abnormalities.

The thyroid gland consists of numerous spherical follicles composed of thyroid follicular cells that surround secreted colloid, a proteinaceous fluid containing large amounts of thyroglobulin, the protein precursor of thyroid hormones (**Fig. 375-2**). The thyroid follicular cells are polarized—the basolateral surface is apposed to the bloodstream and an apical surface faces the follicular lumen. Increased demand for thyroid hormone is regulated by TSH, which binds to its receptor on the basolateral surface of the follicular cells. This binding leads to Tg reabsorption from the follicular lumen and proteolysis within the cytoplasm, yielding thyroid hormones for secretion into the bloodstream.

FIGURE 375-2

Regulation of thyroid hormone synthesis. *Left.* Thyroid hormones T_4 and T_3 feedback to inhibit hypothalamic production of thyrotropin-releasing hormone (TRH) and pituitary production of thyroid-stimulating hormone (TSH). TSH stimulates thyroid gland production of T_4 and T_3 . *Right.* Thyroid follicles are formed by thyroid epithelial cells surrounding proteinaceous colloid, which contains thyroglobulin. Follicular cells, which are polarized, synthesize thyroglobulin and carry out thyroid hormone biosynthesis (see text for details). DIT, diiodotyrosine; MIT, monoiodotyrosine; NIS, sodium iodide symporter; Tg, thyroglobulin; TPO, thyroid peroxidase; TSH-R, thyroid-stimulating hormone receptor.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

REGULATION OF THE THYROID AXIS

TSH, secreted by the thyrotrope cells of the anterior pituitary, plays a pivotal role in control of the thyroid axis and serves as the most useful physiologic marker of thyroid hormone action. TSH is a 31-kDa hormone composed of α and β subunits; the α subunit is common to the other glycoprotein hormones (luteinizing hormone, follicle-stimulating hormone, human chorionic gonadotropin [hCG]), whereas the TSH β subunit is unique to TSH. The extent and nature of carbohydrate modification are modulated by thyrotropin-releasing hormone (TRH) stimulation and influence the biologic activity of the hormone.

The thyroid axis is a classic example of an endocrine feedback loop (Chap. 370). Hypothalamic TRH stimulates pituitary production of TSH, which, in turn, stimulates thyroid hormone synthesis and secretion. Thyroid hormones act via negative feedback predominantly through thyroid hormone receptor $\beta 2$ (TR $\beta 2$) to inhibit TRH and TSH production (Fig. 375-2). The “set-point” in this axis is established by TSH. TRH is the major positive regulator of TSH synthesis and secretion. Peak TSH secretion occurs ~15 min after administration of exogenous TRH. Dopamine, glucocorticoids, and somatostatin suppress TSH but are not of major physiologic importance except when these agents are administered in pharmacologic doses. Reduced levels of thyroid hormone increase basal TSH production and enhance TRH-mediated stimulation of TSH. High thyroid hormone levels rapidly and directly suppress TSH gene expression secretion and inhibit TRH stimulation of TSH, indicating that thyroid hormones are the dominant regulator of TSH production. Like other pituitary hormones, TSH is released in a pulsatile manner and exhibits a diurnal rhythm; its highest levels occur at night. However, these TSH excursions are modest in comparison to those of other pituitary hormones, in part, because TSH has a relatively long plasma half-life (50 min). Consequently, single measurements of TSH are adequate for assessing its circulating level. TSH is measured using immunoradiometric assays that are highly sensitive and specific. These assays readily distinguish between normal and suppressed TSH values; thus, TSH can be used for the diagnosis of primary hyperthyroidism (low TSH) or primary hypothyroidism (high TSH).

THYROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

THYROID HORMONE SYNTHESIS

Thyroid hormones are derived from Tg, a large iodinated glycoprotein. After secretion into the thyroid follicle, Tg is iodinated on tyrosine residues that are subsequently coupled via an ether linkage. Reuptake of Tg into the thyroid follicular cell allows proteolysis and the release of newly synthesized T₄ and T₃.

Iodine Metabolism and Transport

Iodide uptake is a critical first step in thyroid hormone synthesis. Ingested iodine is bound to serum proteins, particularly albumin. Unbound iodine is excreted in the urine. The thyroid gland extracts iodine from the circulation in a highly efficient manner. For example, 10–25% of radioactive tracer (e.g., ¹²³I) is taken up by the normal thyroid gland over 24 h in an iodine-replete state; this value can rise to 70–90% in Graves' disease. Iodide uptake is mediated by NIS, which is expressed at the basolateral membrane of thyroid follicular cells. NIS is most highly expressed in the thyroid gland, but low levels are present in the salivary glands, lactating

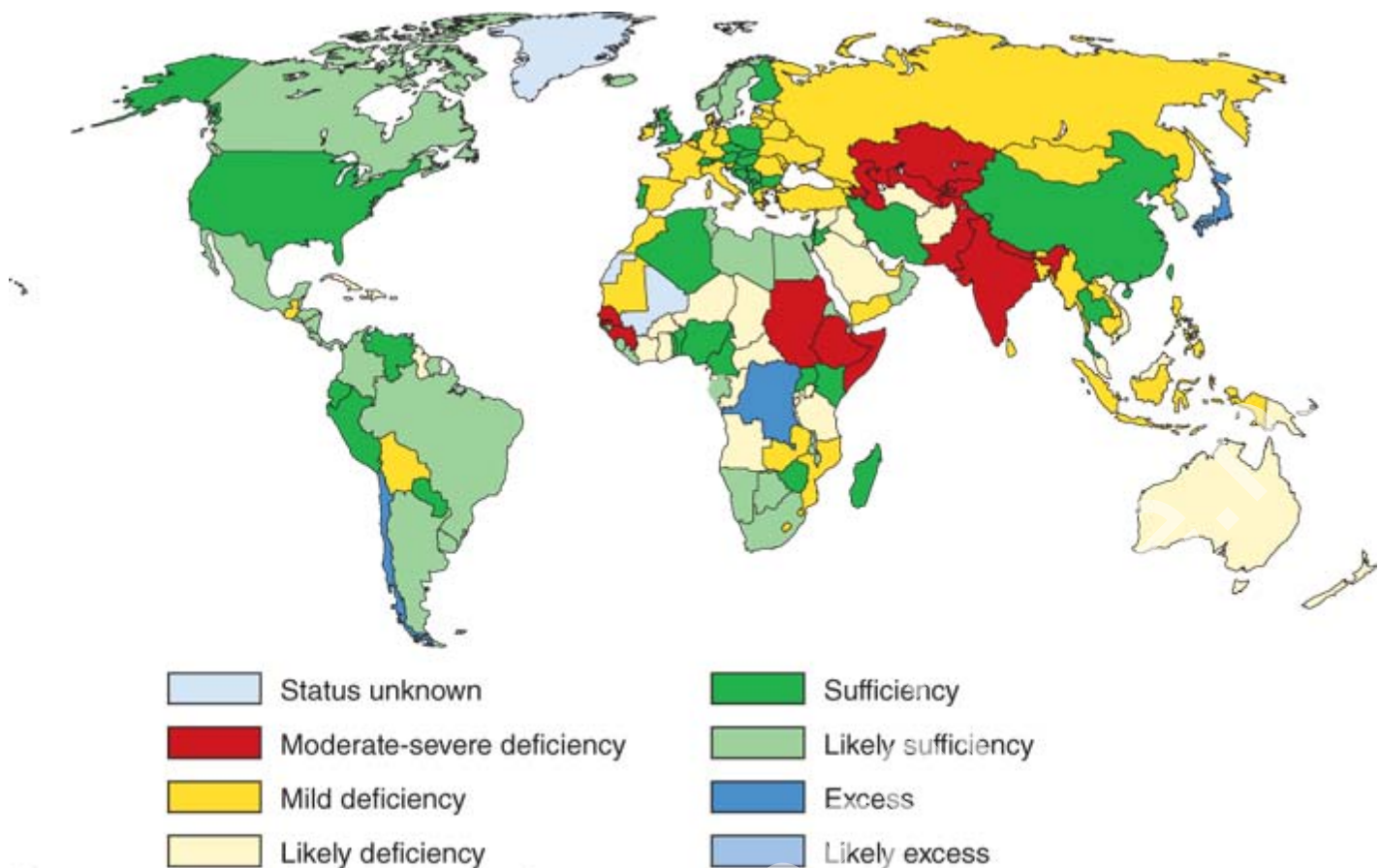
breast, and placenta. The iodide transport mechanism is highly regulated, allowing adaptation to variations in dietary supply. Low iodine levels increase the amount of NIS and stimulate uptake, whereas high iodine levels suppress NIS expression and uptake. The selective expression of NIS in the thyroid allows isotopic scanning, treatment of hyperthyroidism, and ablation of thyroid cancer with radioisotopes of iodine, without significant effects on other organs. Mutation of the *NIS* gene is a rare cause of congenital hypothyroidism, underscoring its importance in thyroid hormone synthesis. Another iodine transporter, pendrin, is located on the apical surface of thyroid cells and mediates iodine efflux into the lumen. Mutation of the *pendrin* gene causes *Pendred syndrome*, a disorder characterized by defective organification of iodine, goiter, and sensorineural deafness.



Iodine deficiency is prevalent in many mountainous regions and in central Africa, central South America, and northern Asia (**Fig. 375-3**). Europe remains mildly iodine-deficient, and health surveys indicate that iodine intake has been falling in the United States and Australia. The World Health Organization (WHO) estimates that about 2 billion people are iodine-deficient, based on urinary excretion data. In areas of relative iodine deficiency, there is an increased prevalence of goiter and, when deficiency is severe, hypothyroidism and cretinism. *Cretinism* is characterized by mental and growth retardation and occurs when children who live in iodine-deficient regions are not treated with iodine or thyroid hormone to restore normal thyroid hormone levels during early life. These children are often born to mothers with iodine deficiency, and it is likely that maternal thyroid hormone deficiency worsens the condition. Concomitant selenium deficiency may also contribute to the neurologic manifestations of cretinism. Iodine supplementation of salt, bread, and other food substances has markedly reduced the prevalence of cretinism. Unfortunately, however, iodine deficiency remains the most common cause of preventable mental deficiency, often because of societal resistance to food additives or the cost of supplementation. In addition to overt cretinism, mild iodine deficiency can lead to subtle reduction of IQ. Oversupply of iodine, through supplements or foods enriched in iodine (e.g., shellfish, kelp), is associated with an increased incidence of autoimmune thyroid disease. The RDA is 220 µg iodine per day for pregnant women and 290 µg iodine per day for breastfeeding women. Because the effects of iodine deficiency are most severe in pregnant women and their babies, the American Thyroid Association has recommended that all pregnant and breastfeeding women in the United States and Canada take a prenatal multivitamin containing 150 µg iodine per day. Urinary iodine is >10 µg/dL in iodine-sufficient populations.

FIGURE 375-3

Worldwide iodine nutrition. Data are from the World Health Organization and the International Council for the Control of Iodine Deficiency Disorders (<http://indorgs.virginia.edu/iccidd/mi/cidds.html>).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Organification, Coupling, Storage, and Release

After iodide enters the thyroid, it is trapped and transported to the apical membrane of thyroid follicular cells, where it is oxidized in an organification reaction that involves TPO and hydrogen peroxide produced by dual oxidase (DUOX) and DUOX maturation factor (DUOXA). The reactive iodine atom is added to specific tyrosyl residues within Tg, a large (660 kDa) dimeric protein that consists of 2769 amino acids. The iodotyrosines in Tg are then coupled via an ether linkage in a reaction that is also catalyzed by TPO. Either T_4 or T_3 can be produced by this reaction, depending on the number of iodine atoms present in the iodotyrosines. After coupling, Tg is taken back into the thyroid cell, where it is processed in lysosomes to release T_4 and T_3 . Uncoupled mono- and diiodotyrosines (MIT, DIT) are deiodinated by the enzyme dehalogenase, thereby recycling any iodide that is not converted into thyroid hormones.

Disorders of thyroid hormone synthesis are rare causes of congenital hypothyroidism ([Chap. 376](#)). The vast majority of these disorders are due to recessive mutations in TPO or Tg, but defects have also been identified in the TSH-R, NIS, pendrin, hydrogen peroxide generation, and dehalogenase, as well as genes involved in thyroid gland development. Because of the biosynthetic defect, the gland is incapable of synthesizing adequate amounts of hormone, leading to increased TSH and a large goiter.

TSH Action

TSH regulates thyroid gland function through the TSH-R, a seven-transmembrane G protein–coupled receptor (GPCR). The TSH-R is coupled to the α subunit of stimulatory G protein ($G_{s\alpha}$), which activates adenylyl cyclase, leading to increased production of cyclic adenosine monophosphate (AMP). TSH also stimulates phosphatidylinositol turnover by activating phospholipase C. The functional role of the TSH-R is exemplified by the consequences of naturally occurring mutations. Recessive loss-of-function mutations cause thyroid hypoplasia and congenital hypothyroidism. Dominant gain-of-function mutations cause sporadic or familial hyperthyroidism that is characterized by goiter, thyroid cell hyperplasia, and autonomous function ([Chap. 377](#)). Most of these activating mutations occur in the transmembrane domain of the receptor. They mimic the conformational changes induced by TSH binding or the interactions of thyroid-stimulating immunoglobulins (TSI) in Graves' disease. Activating TSH-R mutations also occur as somatic events, leading to clonal selection and expansion of the affected thyroid follicular cell and autonomously functioning thyroid nodules.

Other Factors That Influence Hormone Synthesis and Release

Although TSH is the dominant hormonal regulator of thyroid gland growth and function, a variety of growth factors, most produced locally in the thyroid gland, also influence thyroid hormone synthesis. These include insulin-like growth factor I (IGF-I), epidermal growth factor, transforming growth factor β (TGF- β), endothelins, and various cytokines. The quantitative roles of these factors are not well understood, but they are important in selected disease states. In acromegaly, for example, increased levels of growth hormone and IGF-I are associated with goiter and predisposition to multinodular goiter (MNG). Certain cytokines and interleukins (ILs) produced in association with autoimmune thyroid disease induce thyroid growth, whereas others lead to apoptosis. Iodine deficiency increases thyroid blood flow and upregulates the NIS, stimulating more efficient iodine uptake. Excess iodide transiently inhibits thyroid iodide organification, a phenomenon known as the *Wolff-Chaikoff effect*. In individuals with a normal thyroid, the gland escapes from this inhibitory effect and iodide organification resumes; the suppressive action of high iodide may persist, however, in patients with underlying autoimmune thyroid disease.

THYROID FUNCTION IN PREGNANCY

Five factors alter thyroid function in pregnancy: (1) the transient increase in hCG during the first trimester, which weakly stimulates the TSH-R; (2) the estrogen-induced rise in TBG during the first trimester, which is sustained during pregnancy; (3) alterations in the immune system, leading to the onset, exacerbation, or amelioration of an underlying autoimmune thyroid disease; (4) increased thyroid hormone metabolism by the placenta; and (5) increased urinary iodide excretion, which can cause impaired thyroid hormone production in areas of marginal iodine sufficiency. Women with a precarious iodine intake ($<50 \mu\text{g/d}$) are most at risk of developing a goiter during pregnancy or giving birth to an infant with a goiter and hypothyroidism. The World Health Organization recommends a daily iodine intake of $250 \mu\text{g}$ during pregnancy and lactation and prenatal vitamins should contain $150 \mu\text{g}$ per tablet.

The rise in circulating hCG levels during the first trimester is accompanied by a reciprocal fall in TSH that persists into the middle of pregnancy. This reflects the weak binding of hCG, which is present at very high levels, to the TSH-R. Rare individuals have variant TSH-R sequences that enhance hCG binding and TSH-R activation. hCG-induced changes in thyroid function can result in transient gestational hyperthyroidism that may be associated with *hyperemesis gravidarum*, a condition characterized by severe nausea and vomiting and risk of volume depletion. However, since the hyperthyroidism is not causal, antithyroid drugs are not indicated unless concomitant Graves' disease is suspected. Parenteral fluid replacement usually suffices until the condition resolves.

Normative values for most thyroid function tests differ during pregnancy and, if available, trimester specific reference ranges should be used when diagnosing thyroid dysfunction during pregnancy. TSH levels decrease during the first trimester and then rise as gestation progresses. Total T_4 and T_3 levels are about 1.5× higher throughout pregnancy but the free T_4 progressively decreases so that third trimester values in healthy pregnancies are often below the nonpregnant lower reference cutoff.

During pregnancy, subclinical hypothyroidism occurs in 2% of women, but overt hypothyroidism is present in only 1 in 500. Prospective randomized controlled trials have not shown a benefit for universal thyroid disease screening in pregnancy. Targeted TSH testing for hypothyroidism is recommended for women planning a pregnancy if they have a strong family history of autoimmune thyroid disease, other autoimmune disorders (e.g., type 1 diabetes), infertility, prior preterm delivery or recurrent miscarriage, signs or symptoms of thyroid disease, or are older than 30 years. Thyroid hormone requirements are increased by up to 45% during pregnancy in levothyroxine-treated hypothyroid women.

THYROID HORMONE TRANSPORT AND METABOLISM

Serum-Binding Proteins

T_4 is secreted from the thyroid gland in about twentyfold excess over T_3 (Table 375-1). Both hormones are bound to plasma proteins, including thyroxine-binding globulin (TBG), transthyretin (TTR, formerly known as thyroxine-binding prealbumin, or TBPA), and albumin. The plasma-binding proteins increase the pool of circulating hormone, delay hormone clearance, and may modulate hormone delivery to selected tissue sites. The concentration of TBG is relatively low (1–2 mg/dL), but because of its high affinity for thyroid hormones ($T_4 > T_3$), it carries about 80% of the bound hormones. Albumin has relatively low affinity for thyroid hormones but has a high plasma concentration (~3.5 g/dL), and it binds up to 10% of T_4 and 30% of T_3 . TTR carries about 10% of T_4 but little T_3 .

TABLE 375-1

Characteristics of Circulating T₄ and T₃

HORMONE PROPERTY	T ₄	T ₃
Serum concentrations		
Total hormone	8 µg/dL	0.14 µg/dL
Fraction of total hormone in the unbound form	0.02%	0.3%
Unbound (free) hormone	$21 \times 10^{-12} M$	$6 \times 10^{-12} M$
Serum half-life	7 d	2 d
Fraction directly from the thyroid	100%	20%
Production rate, including peripheral conversion	90 µg/d	32 µg/d
Intracellular hormone fraction	~20%	~70%
Relative metabolic potency	0.3	1
Receptor binding	$10^{-10} M$	$10^{-11} M$

When the effects of the various binding proteins are combined, ~99.98% of T₄ and 99.7% of T₃ are protein-bound. Because T₃ is less tightly bound than T₄, the fraction of unbound T₃ is greater than unbound T₄, but there is less unbound T₃ in the circulation because it is produced in smaller amounts and cleared more rapidly than T₄. The unbound or “free” concentrations of the hormones are $\sim 2 \times 10^{-11} M$ for T₄ and $\sim 6 \times 10^{-12} M$ for T₃, which roughly correspond to the thyroid hormone receptor-binding constants for these hormones (see below). The unbound hormone is thought to be biologically available to tissues. Nonetheless, the homeostatic mechanisms that regulate the thyroid axis are directed toward maintenance of normal concentrations of unbound hormones.

Abnormalities of Thyroid Hormone-Binding Proteins

A number of inherited and acquired abnormalities affect thyroid hormone-binding proteins. X-linked TBG deficiency is associated with very low levels of total T₄ and T₃. However, because unbound hormone levels

are normal, patients are euthyroid and TSH levels are normal. It is important to recognize this disorder to avoid efforts to normalize total T_4 levels, because this leads to thyrotoxicosis and is futile because of rapid hormone clearance in the absence of TBG. TBG levels are elevated by estrogen, which increases sialylation and delays TBG clearance. Consequently, in women who are pregnant or taking estrogen-containing contraceptives, elevated TBG increases total T_4 and T_3 levels; however, unbound T_4 and T_3 levels are normal. These features are part of the explanation for why women with hypothyroidism require increased amounts of L-thyroxine replacement as TBG levels are increased by pregnancy or estrogen treatment. Mutations in TBG, TTR, and albumin may increase the binding affinity for T_4 and/or T_3 and cause disorders known as *euthyroid hyperthyroxinemia* or *familial dysalbuminemic hyperthyroxinemia* (FDH) (Table 375-2). These disorders result in increased total T_4 and/or T_3 , but unbound hormone levels are normal. The familial nature of the disorders, and the fact that TSH levels are normal rather than suppressed, should suggest this diagnosis. Unbound hormone levels (ideally measured by dialysis) are normal in FDH. The diagnosis can be confirmed by using tests that measure the affinities of radiolabeled hormone binding to specific transport proteins or by performing DNA sequence analyses of the abnormal transport protein genes.

TABLE 375-2

Conditions Associated with Euthyroid Hyperthyroxinemia

DISORDER	CAUSE	TRANSMISSION	CHARACTERISTICS
Familial dysalbuminemic hyperthyroxinemia (FDH)	Albumin mutations, usually R218H	AD	Increased T ₄ Normal unbound T ₄ Rarely increased T ₃
TBG			
Familial excess	Increased TBG production	XL	Increased total T ₄ , T ₃ Normal unbound T ₄ , T ₃
Acquired excess	Medications (estrogen), pregnancy, cirrhosis, hepatitis	Acquired	Increased total T ₄ , T ₃ Normal unbound T ₄ , T ₃
Transthyretin ^a			
Excess	Islet tumors	Acquired	Usually normal T ₄ , T ₃
Mutations	Increased affinity for T ₄ or T ₃	AD	Increased total T ₄ , T ₃ Normal unbound T ₄ , T ₃
Medications: propranolol, ipodate, iopanoic acid, amiodarone	Decreased T ₄ → T ₃ conversion	Acquired	Increased T ₄ Decreased T ₃ Normal or increased TSH

DISORDER	CAUSE	TRANSMISSION	CHARACTERISTICS
Resistance to thyroid hormone (RTH)	Thyroid hormone receptor β mutations	AD	Increased unbound T_4 , T_3 Normal or increased TSH Some patients clinically thyrotoxic

^aAlso known as thyroxine-binding prealbumin (TBPA).

Abbreviations: AD, autosomal dominant; TBG, thyroxine-binding globulin; TSH, thyroid-stimulating hormone; XL, X-linked.

Certain medications, such as salicylates and salsalate, can displace thyroid hormones from circulating binding proteins. Although these drugs transiently perturb the thyroid axis by increasing free thyroid hormone levels, TSH is suppressed until a new steady state is reached, thereby restoring euthyroidism. Circulating factors associated with acute illness may also displace thyroid hormone from binding proteins (Chap. 377).

Deiodinases

T_4 may be thought of as a precursor for the more potent T_3 . T_4 is converted to T_3 by the deiodinase enzymes (Fig. 375-1). Type I deiodinase, which is located primarily in thyroid, liver, and kidneys, has a relatively low affinity for T_4 . Type II deiodinase has a higher affinity for T_4 and is found primarily in the pituitary gland, brain, brown fat, and thyroid gland. Expression of type II deiodinase allows it to regulate T_3 concentrations locally, a property that may be important in the context of levothyroxine (T_4) replacement. Type II deiodinase is also regulated by thyroid hormone; hypothyroidism induces the enzyme, resulting in enhanced $T_4 \rightarrow T_3$ conversion in tissues such as brain and pituitary. $T_4 \rightarrow T_3$ conversion is impaired by fasting, systemic illness or acute trauma, oral contrast agents, and a variety of medications (e.g., propylthiouracil, propranolol, amiodarone, glucocorticoids). Type III deiodinase inactivates T_4 and T_3 and is the most important source of reverse T_3 (rT_3), including in the sick euthyroid syndrome. This enzyme is expressed in the human placenta but is not active in healthy individuals. In the sick euthyroid syndrome, especially with hypoperfusion, the type III deiodinase is activated in muscle and liver. Massive hemangiomas that express type III deiodinase are a rare cause of consumptive hypothyroidism in infants.

THYROID HORMONE ACTION

Thyroid Hormone Transport

Circulating thyroid hormones enter cells by passive diffusion and via specific transporters such as the monocarboxylate 8 transporter (MCT8), MCT10, and organic anion-transporting polypeptide 1C1. Mutations in the *MCT8* gene have been identified in patients with X-linked psychomotor retardation and thyroid function abnormalities (low T_4 , high T_3 , and high TSH). After entering cells, thyroid hormones act primarily through nuclear receptors, although they also have nongenomic actions through stimulating mitochondrial enzymatic responses and may act directly on blood vessels and the heart through integrin receptors.

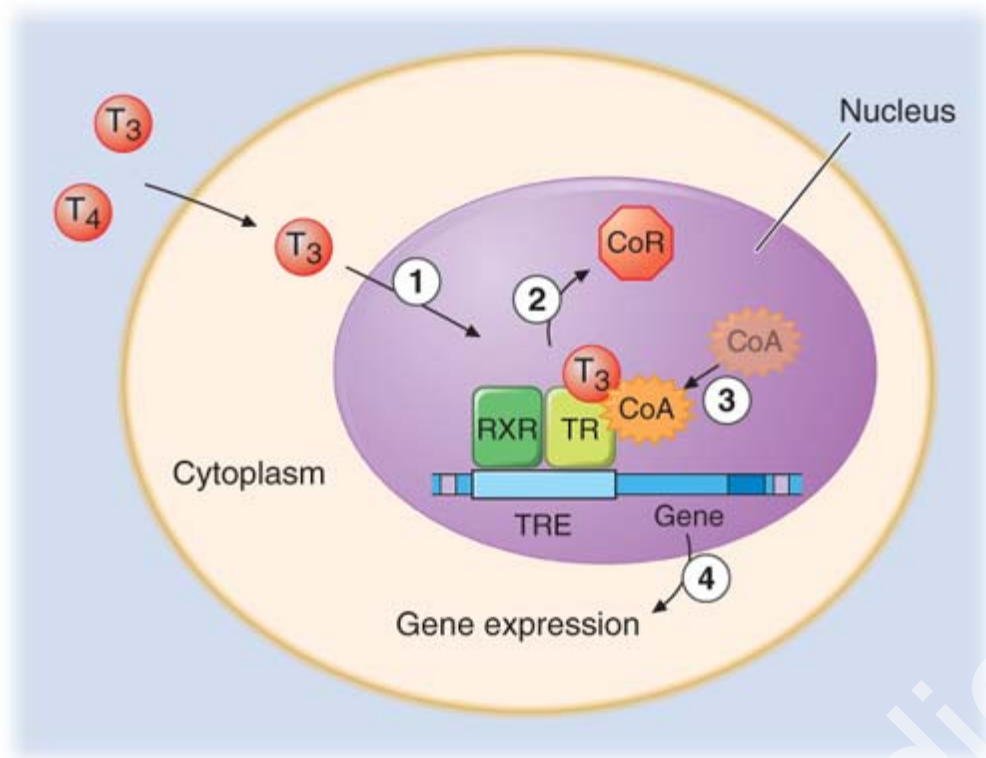
Nuclear Thyroid Hormone Receptors

Thyroid hormones bind with high affinity to nuclear *thyroid hormone receptors* (TRs) α and β . Both TR α and TR β are expressed in most tissues, but their relative expression levels vary among organs; TR α is particularly abundant in brain, kidneys, gonads, muscle, and heart, whereas TR β expression is relatively high in the pituitary and liver. Both receptors are variably spliced to form unique isoforms. The TR β 2 isoform, which has a unique amino terminus, is selectively expressed in the hypothalamus and pituitary, where it plays a role in feedback control of the thyroid axis (see above). The TR α 2 isoform contains a unique carboxy terminus that precludes thyroid hormone binding; it may function to inhibit the action of other TR isoforms.

The TRs contain a central DNA-binding domain and a C-terminal ligand-binding domain. They bind to specific DNA sequences, termed *thyroid response elements* (TREs), in the promoter regions of target genes (**Fig. 375-4**). The receptors bind as homodimers or, more commonly, as heterodimers with retinoic acid X receptors (RXRs) (**Chap. 370**). The activated receptor can either stimulate gene transcription (e.g., myosin heavy chain α) or inhibit transcription (e.g., TSH β -subunit gene), depending on the nature of the regulatory elements in the target gene.

FIGURE 375-4

Mechanism of thyroid hormone receptor action. The thyroid hormone receptor (TR) and retinoid X receptor (RXR) form heterodimers that bind specifically to thyroid hormone response elements (TRE) in the promoter regions of target genes. In the absence of hormone, TR binds co-repressor (CoR) proteins that silence gene expression. The numbers refer to a series of ordered reactions that occur in response to thyroid hormone: (1) T_4 or T_3 enters the nucleus; (2) T_3 binding dissociates CoR from TR; (3) co-activators (CoA) are recruited to the T_3 -bound receptor; and (4) gene expression is altered.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Thyroid hormones (T₃ and T₄) bind with similar affinities to TR α and TR β . However, structural differences in the ligand-binding domains provide the potential for developing receptor-selective agonists or antagonists, and these are under investigation. T₃ is bound with 10–15 times greater affinity than T₄, which explains its increased hormonal potency. Although T₄ is produced in excess of T₃, receptors are occupied mainly by T₃, reflecting T₄ \rightarrow T₃ conversion by peripheral tissues, T₃ bioavailability in the plasma, and the greater affinity of receptors for T₃. After binding to TRs, thyroid hormone induces conformational changes in the receptors that modify its interactions with accessory transcription factors. Importantly, in the absence of thyroid hormone binding, the aporeceptors bind to co-repressor proteins that inhibit gene transcription. Hormone binding dissociates the co-repressors and allows the recruitment of co-activators that enhance transcription. The discovery of TR interactions with co-repressors explains the fact that TR silences gene expression in the absence of hormone binding. Consequently, hormone deficiency has a profound effect on gene expression because it causes gene repression as well as loss of hormone-induced stimulation. This concept has been corroborated by the finding that targeted deletion of the TR genes in mice has a less pronounced phenotypic effect than hormone deficiency.

Thyroid Hormone Resistance

Resistance to thyroid hormone (RTH) is an autosomal dominant disorder characterized by elevated thyroid hormone levels and inappropriately normal or elevated TSH. Individuals with RTH do not, in general, exhibit signs and symptoms that are typical of hypothyroidism because hormone resistance is partial and is compensated by increased levels of thyroid hormone. The clinical features of RTH can include goiter,

attention deficit disorder, mild reduction in IQ, delayed skeletal maturation, tachycardia, and impaired metabolic responses to thyroid hormone.

Classical forms of RTH are caused by mutations in the TR β gene. These mutations, located in restricted regions of the ligand-binding domain, cause loss of receptor function. However, because the mutant receptors retain the capacity to dimerize with RXRs, bind to DNA, and recruit co-repressor proteins, they function as antagonists of the remaining normal TR β and TR α receptors. This property, referred to as “dominant negative” activity, explains the autosomal dominant mode of transmission. The diagnosis is suspected when unbound thyroid hormone levels are increased without suppression of TSH. Similar hormonal abnormalities are found in other affected family members, although the TR β mutation arises de novo in about 20% of patients. DNA sequence analysis of the TR β gene provides a definitive diagnosis. RTH must be distinguished from other causes of euthyroid hyperthyroxinemia (e.g., FDH) and inappropriate secretion of TSH by TSH-secreting pituitary adenomas ([Chap. 373](#)). In most patients, no treatment is indicated; the importance of making the diagnosis is to avoid inappropriate treatment of mistaken hyperthyroidism and to provide genetic counseling.

A distinct form of RTH is caused by mutations in the TR α gene. Affected patients have many clinical features of congenital hypothyroidism including growth retardation, skeletal dysplasia, and severe constipation. In contrast to RTH caused by mutations in TR β , thyroid function tests include normal TSH, low or normal T₄, and normal or elevated T₃ levels. These distinct clinical and laboratory features underscore the different tissue distribution and functional roles of TR β and TR α . Thyroxine treatment appears to alleviate some of the clinical manifestations of patients with RTH caused by TR α mutations.

PHYSICAL EXAMINATION

In addition to the examination of the thyroid itself, the physical examination should include a search for signs of abnormal thyroid function and the extrathyroidal features of ophthalmopathy and dermopathy ([Chap. 377](#)). Examination of the neck begins by inspecting the seated patient from the front and side and noting any surgical scars, obvious masses, or distended veins. The thyroid can be palpated with both hands from behind or while facing the patient, using the thumbs to palpate each lobe. It is best to use a combination of these methods, especially when nodules are small. The patient’s neck should be slightly flexed to relax the neck muscles. After locating the cricoid cartilage, the isthmus, which is attached to the lower one-third of the thyroid lobes, can be identified and then followed laterally to locate either lobe (normally, the right lobe is slightly larger than the left). By asking the patient to swallow sips of water, thyroid consistency can be better appreciated as the gland moves beneath the examiner’s fingers.

Features to be noted include thyroid size, consistency, nodularity, and any tenderness or fixation. An estimate of thyroid size (normally 12–20 g) should be made, and a drawing is often the best way to record findings. Ultrasound imaging provides the most accurate measurement of thyroid volume and nodularity and is useful for assessment of goiter prevalence in iodine deficient regions. However, ultrasound is not indicated if the thyroid physical examination is normal. The size, location, and consistency of any nodules

should also be defined. A bruit or thrill over the gland, located over the insertion of the superior and inferior thyroid arteries (supero- or inferolaterally), indicates increased vascularity, associated with turbulent rather than laminar blood flow, as occurs in hyperthyroidism. If the lower borders of the thyroid lobes are not clearly felt, a goiter may be retrosternal. Large retrosternal goiters can cause venous distention over the neck and difficulty breathing, especially when the arms are raised (Pemberton's sign). With any central mass above the thyroid, the tongue should be extended, as thyroglossal cysts then move upward. The thyroid examination is not complete without assessment for lymphadenopathy in the supraclavicular and cervical regions of the neck.

LABORATORY EVALUATION

Measurement of Thyroid Hormones

The enhanced sensitivity and specificity of *TSH assays* have greatly improved laboratory assessment of thyroid function. Because TSH levels change dynamically in response to alterations of T_4 and T_3 , a logical approach to thyroid testing is to first determine whether TSH is suppressed, normal, or elevated. With rare exceptions (see below), a normal TSH level excludes a primary abnormality of thyroid function. This strategy depends on the use of immunochemiluminometric assays (ICMAs) for TSH that are sensitive enough to discriminate between the lower limit of the reference interval and the suppressed values that occur with thyrotoxicosis. Extremely sensitive assays can detect TSH levels ≤ 0.004 mIU/L, but, for practical purposes, assays sensitive to ≤ 0.1 mIU/L are sufficient. The widespread availability of the TSH ICMA has rendered the TRH stimulation test obsolete, because the failure of TSH to rise after an intravenous bolus of 200–400 μ g TRH has the same implications as a suppressed basal TSH measured by ICMA.

The finding of an abnormal TSH level must be followed by measurements of circulating thyroid hormone levels to confirm the diagnosis of hyperthyroidism (suppressed TSH) or hypothyroidism (elevated TSH). Automated immunoassays are widely available for serum *total* T_4 and *total* T_3 . T_4 and T_3 are highly protein-bound, and numerous factors (illness, medications, genetic factors) can influence protein binding. It is useful, therefore, to measure the free, or unbound, hormone levels, which correspond to the biologically available hormone pool. Two direct methods are used to measure *unbound thyroid hormones*: (1) unbound thyroid hormone competition with radiolabeled T_4 (or an analogue) for binding to a solid-phase antibody, and (2) physical separation of the unbound hormone fraction by ultracentrifugation or equilibrium dialysis. Although early unbound hormone immunoassays suffered from artifacts, newer assays correlate well with the results of the more technically demanding and expensive physical separation methods. An indirect method that is now less commonly used to estimate unbound thyroid hormone levels is to calculate the free T_3 or free T_4 index from the total T_4 or T_3 concentration and the *thyroid hormone binding ratio* (THBR). The latter is derived from the *T_3 -resin uptake test*, which determines the distribution of radiolabeled T_3 between an absorbent resin and the unoccupied thyroid hormone binding proteins in the sample. The binding of the labeled T_3 to the resin is increased when there is reduced unoccupied protein binding sites (e.g., TBG deficiency) or increased total thyroid hormone in the sample; it is decreased under the opposite

circumstances. The product of THBR and total T_3 or T_4 provides the *free T_3 or T_4 index*. In effect, the index corrects for anomalous total hormone values caused by variations in hormone-protein binding.

Total thyroid hormone levels are *elevated* when TBG is increased due to **estrogens** (pregnancy, oral contraceptives, hormone therapy, tamoxifen, selective estrogen receptor modulators, inflammatory liver disease) and *decreased* when TBG binding is reduced (androgens, nephrotic syndrome). Genetic disorders and acute illness can also cause abnormalities in thyroid hormone-binding proteins, and various drugs (phenytoin, carbamazepine, salicylates, and nonsteroidal anti-inflammatory drugs [NSAIDs]) can interfere with thyroid hormone binding. Because unbound thyroid hormone levels are normal and the patient is euthyroid in all of these circumstances, assays that measure unbound hormone are preferable to those for total thyroid hormones.

For most purposes, the unbound T_4 level is sufficient to confirm thyrotoxicosis, but 2–5% of patients have only an elevated T_3 level (T_3 toxicosis). Thus, unbound T_3 levels should be measured in patients with a suppressed TSH but normal unbound T_4 levels.

There are several clinical conditions in which the use of TSH as a screening test may be misleading, particularly without simultaneous unbound T_4 determinations. Any severe nonthyroidal illness can cause abnormal TSH levels. Although hypothyroidism is the most common cause of an elevated TSH level, rare causes include a TSH-secreting pituitary tumor (**Chap. 373**), thyroid hormone resistance, and assay artifact. Conversely, a suppressed TSH level, particularly <0.01 mIU/L, usually indicates thyrotoxicosis. However, subnormal TSH levels between 0.01 and 0.1 mIU/L may be seen during the first trimester of pregnancy (due to hCG secretion), after treatment of hyperthyroidism (because TSH can remain suppressed for several months), and in response to certain medications (e.g., high doses of glucocorticoids or dopamine). TSH levels measured by immunoassay may also be suppressed in patients ingesting biotin supplements <18 hours prior to a blood draw because the TSH capture antibodies are biotinylated and the exogenous biotin can interfere with the subsequent streptavidin capture. Importantly, secondary hypothyroidism, caused by hypothalamic-pituitary disease, is associated with a variable (low to high-normal) TSH level, which is inappropriate for the low T_4 level. Thus, *TSH should not be used as an isolated laboratory test to assess thyroid function in patients with suspected or known pituitary disease.*

Tests for the end-organ effects of thyroid hormone excess or depletion, such as estimation of basal metabolic rate, tendon reflex relaxation rates, or serum cholesterol, are relatively insensitive and are not useful as clinical determinants of thyroid function.

Tests to Determine the Etiology of Thyroid Dysfunction

Autoimmune thyroid disease is detected most easily by measuring circulating antibodies against TPO and Tg. Because antibodies to Tg alone are uncommon, it is reasonable to measure only TPO antibodies. About 5–15% of euthyroid women and up to 2% of euthyroid men have thyroid antibodies; such individuals are at

increased risk of developing thyroid dysfunction. Almost all patients with autoimmune hypothyroidism, and up to 80% of those with Graves' disease, have TPO antibodies, usually at high levels.

TSIs are antibodies that stimulate the TSH-R in Graves' disease. They are most commonly measured by commercially available tracer displacement assays called TRAb (TSH receptor antibody) with the assumption that elevated levels in the setting of clinical hyperthyroidism reflect stimulatory effects on the TSH receptor. A bioassay is less commonly used. Remission rates in patients with Graves' disease after antithyroid drug cessation are higher with disappearance rather than persistence of TRAb. Furthermore, the TRAb assay is used to predict both fetal and neonatal thyrotoxicosis caused by transplacental passage of high maternal levels of TRAb or TSI ($>3\times$ upper limit of normal) in the last trimester of pregnancy.

Serum Tg levels are increased in all types of thyrotoxicosis except *thyrotoxicosis factitia* caused by self-administration of thyroid hormone. Tg levels are particularly increased in thyroiditis, reflecting thyroid tissue destruction and release of Tg. The main role for Tg measurement, however, is in the follow-up of thyroid cancer patients. After total thyroidectomy and radioablation, Tg levels should be undetectable; in the absence of anti-Tg antibodies, measurable levels indicate incomplete ablation or recurrent cancer.

Radioiodine Uptake and Thyroid Scanning

The thyroid gland selectively transports radioisotopes of iodine (^{123}I , ^{125}I , ^{131}I) and $^{99\text{m}}\text{Tc}$ pertechnetate, allowing thyroid imaging and quantitation of radioactive tracer fractional uptake.

Nuclear imaging of Graves' disease is characterized by an enlarged gland and increased tracer uptake that is distributed homogeneously. Toxic adenomas appear as focal areas of increased uptake, with suppressed tracer uptake in the remainder of the gland. In toxic MNG, the gland is enlarged—often with distorted architecture—and there are multiple areas of relatively increased (functioning nodules) or decreased tracer uptake (suppressed thyroid parenchyma or nonfunctioning nodules). Subacute, viral, and postpartum thyroiditis are associated with very low uptake because of follicular cell damage and TSH suppression. Thyrotoxicosis factitia is also associated with low uptake. In addition, if there is excessive circulating exogenous iodine (e.g., from dietary sources of iodinated contrast dye), the radionuclide uptake is low even in the presence of increased thyroid hormone production.

Thyroid scintigraphy is not used in the routine evaluation of patients with thyroid nodules, but should be performed if the serum TSH level is subnormal to determine if functioning thyroid nodules are present. Functioning or "hot" nodules are almost never malignant, and fine-needle aspiration (FNA) biopsy is not indicated. The vast majority of thyroid nodules do not produce thyroid hormone ("cold" nodules), and these are more likely to be malignant (~5–10%). Whole-body and thyroid scanning is also used in the treatment and surveillance of thyroid cancer. After thyroidectomy for thyroid cancer, the TSH level is raised by either using a thyroid hormone withdrawal protocol or recombinant human TSH injection ([Chap. 378](#)).

Administration of either ^{131}I or ^{123}I (in higher activities than used to image the thyroid gland alone) allows whole-body scanning (WBS) to confirm remnant ablation and to detect any functioning metastases. In addition, WBS may be helpful in surveillance of patients at risk for recurrence.

Ultrasonography is valuable for the diagnosis and evaluation of patients with nodular thyroid disease ([Chap. 378](#)). Evidence-based guidelines recommend thyroid ultrasonography for all patients suspected of having thyroid nodules by either physical examination or another imaging study. Using 10- to 12-MHz linear transducers, resolution and image quality are excellent, allowing the characterization of nodules and cysts >3 mm. Sonographic patterns that combine suspicious sonographic features are highly suggestive of malignancy (e.g., hypoechoic solid nodules with infiltrative borders and microcalcifications, >90% cancer risk), whereas other patterns correlate with a lower likelihood of cancer (isoechoic solid nodules, 5–10% cancer risk). Some patterns suggest benignity (e.g., spongiform nodules, defined as those with multiple small internal cystic areas, or simple cysts <3% cancer risk) ([see Chap. 378, Fig. 378-1](#)). In addition to evaluating thyroid nodules, ultrasound is useful for monitoring nodule size and for the aspiration of nodules or cystic lesions. Ultrasound-guided FNA biopsy of thyroid lesions lowers the rate of inadequate sampling and decreases sample error, thereby reducing both the nondiagnostic and false-negative rates of FNA cytology. Ultrasonography of the central and lateral cervical lymph node compartments is indispensable in the evaluation thyroid cancer patients, preoperatively and during follow-up. In addition, the American College of Radiology recommends a survey of the cervical lymph nodes as part of every diagnostic thyroid sonographic examination.

FURTHER READING

Alexander EK et al: 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* 27:315, 2017.

Ortiga-Cavalho TM, Sidhaye AR, Wondisford FE: Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat Rev Endocrinol* 10:582, 2014. [[PubMed: 25135573](#)]

Rugge JB, Bougatsos C, Chou R: Screening and treatment of thyroid dysfunction: An evidence review for the U. S. Preventive Services Task Force. *Ann Intern Med* 162:35, 2015. [[PubMed: 25347444](#)]

Stoupa A et al: Update of thyroid developmental genes. *Endocrinol Metab Clin North Am* 45:243, 2016. [[PubMed: 27241962](#)]

Zimmermann MB, Boelaert K: Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol* 3: 286, 2015. [[PubMed: 25591468](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Harrison's Principles of Internal Medicine, 20e >

Chapter 376: Hypothyroidism

J. Larry Jameson; Susan J. Mandel; Anthony P. Weetman

HYPOTHYROIDISM

Iodine deficiency remains a common cause of hypothyroidism worldwide. In areas of iodine sufficiency, autoimmune disease (Hashimoto's thyroiditis) and iatrogenic causes (treatment of hyperthyroidism) are most common ([Table 376-1](#)).

TABLE 376-1

Causes of Hypothyroidism

Primary
Autoimmune hypothyroidism: Hashimoto's thyroiditis, atrophic thyroiditis
Iatrogenic: ^{131}I treatment, subtotal or total thyroidectomy, external irradiation of neck for lymphoma or cancer
Drugs: iodine excess (including iodine-containing contrast media and amiodarone), lithium, antithyroid drugs, <i>p</i> -aminosalicylic acid, interferon α and other cytokines, aminoglutethimide, tyrosine kinase inhibitors (e.g., sunitinib)
Congenital hypothyroidism: absent or ectopic thyroid gland, dyshormonogenesis, TSH-R mutation
Iodine deficiency
Infiltrative disorders: amyloidosis, sarcoidosis, hemochromatosis, scleroderma, cystinosis, Riedel's thyroiditis
Overexpression of type 3 deiodinase in infantile hemangioma and other tumors
Transient
Silent thyroiditis, including postpartum thyroiditis
Subacute thyroiditis
Withdrawal of supraphysiologic thyroxine treatment in individuals with an intact thyroid
After ^{131}I treatment or subtotal thyroidectomy for Graves' disease
Secondary
Hypopituitarism: tumors, pituitary surgery or irradiation, infiltrative disorders, Sheehan's syndrome, trauma, genetic forms of combined pituitary hormone deficiencies
Isolated TSH deficiency or inactivity
Bexarotene treatment
Hypothalamic disease: tumors, trauma, infiltrative disorders, idiopathic

Abbreviations: TSH, thyroid-stimulating hormone; TSH-R, TSH receptor.

CONGENITAL HYPOTHYROIDISM

Prevalence

Hypothyroidism occurs in about 1 in 4000 newborns and neonatal screening is performed in most industrialized countries. It may be transient, especially if the mother has TSH-R blocking antibodies or has received antithyroid drugs, but permanent hypothyroidism occurs in the majority. Neonatal hypothyroidism is due to thyroid gland dysgenesis in 80–85%, to inborn errors of thyroid hormone synthesis in 10–15%, and is TSH-R antibody-mediated in 5% of affected newborns. The developmental abnormalities are twice as common in girls. Mutations that cause congenital hypothyroidism are being increasingly identified, but most remain idiopathic ([Table 376-2](#)). Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides partial hormone support to a fetus with congenital hypothyroidism.

TABLE 376-2

Genetic Causes of Congenital Hypothyroidism

DEFECTIVE GENE PROTEIN	INHERITANCE	CONSEQUENCES
PROP-1	Autosomal recessive	Combined pituitary hormone deficiencies with preservation of adrenocorticotrophic hormone
PIT-1	Autosomal recessive Autosomal dominant	Combined deficiencies of growth hormone, prolactin, thyroid-stimulating hormone (TSH)
TSH β	Autosomal recessive	TSH deficiency
TTF-1 (TITF-1)	Autosomal dominant	Variable thyroid hypoplasia, choreoathetosis, pulmonary problems
TTF-2 (FOXE-1)	Autosomal recessive	Thyroid agenesis, choanal atresia, spiky hair
PAX-8	Autosomal dominant	Thyroid dysgenesis, kidney abnormalities
NKX2-1	Autosomal dominant	Thyroid dysgenesis, brain, lung abnormalities
NKX2-5	Autosomal dominant	Thyroid dysgenesis, heart abnormalities
TSH-receptor	Autosomal recessive	Resistance to TSH
G $s\alpha$ (Albright hereditary osteodystrophy)	Autosomal dominant	Resistance to TSH
Na $^{+}$ /I $^{-}$ symporter (SLC5A5)	Autosomal recessive	Inability to transport iodide

DEFECTIVE GENE PROTEIN	INHERITANCE	CONSEQUENCES
DUOX2 (THOX2)	Autosomal dominant	Organification defect
DUOXA2	Autosomal recessive	Organification defect
Thyroid peroxidase	Autosomal recessive	Defective organification of iodide
Thyroglobulin	Autosomal recessive	Defective synthesis of thyroid hormone
Pendrin (SLC26A4)	Autosomal recessive	Pendred syndrome: sensorineural deafness and partial organification defect in thyroid
Dehalogenase 1 (IYD)	Autosomal recessive	Loss of iodide reutilization

Clinical Manifestations

The majority of infants appear normal at birth, and with the use of biochemical screening, few cases are now diagnosed based on clinical features, which include prolonged jaundice, feeding problems, hypotonia, enlarged tongue, delayed bone maturation, and umbilical hernia. Importantly, permanent neurologic damage results if treatment is delayed. Typical features of adult hypothyroidism may also be present ([Table 376-3](#)). Other congenital malformations, especially cardiac, are four times more common in congenital hypothyroidism.

TABLE 376-3

Signs and Symptoms of Hypothyroidism (Descending Order of Frequency)

SYMPTOMS	SIGNS
Tiredness, weakness	Dry coarse skin; cool peripheral extremities
Dry skin	Puffy face, hands, and feet (myxedema)
Feeling cold	Diffuse alopecia
Hair loss	Bradycardia
Difficulty concentrating and poor memory	Peripheral edema
Constipation	Delayed tendon reflex relaxation
Weight gain with poor appetite	Carpal tunnel syndrome
Dyspnea	Serous cavity effusions
Hoarse voice	
Menorrhagia (later oligomenorrhea or amenorrhea)	
Paresthesia	
Impaired hearing	

Diagnosis and Treatment

Because of the severe neurologic consequences of untreated congenital hypothyroidism, neonatal screening programs have been established. These are generally based on measurement of TSH or T_4 levels in heel-prick blood specimens. When the diagnosis is confirmed, T_4 is instituted at a dose of 10–15 $\mu\text{g}/\text{kg}$ per day, and the dose is adjusted by close monitoring of TSH levels. T_4 requirements are relatively great during the first year of life, and a high circulating T_4 level is usually needed to normalize TSH. Early treatment with T_4 results in normal IQ levels, but subtle neurodevelopmental abnormalities may occur in those with the most severe hypothyroidism at diagnosis or when treatment is delayed or suboptimal. If transient hypothyroidism is suspected, or the diagnosis is unclear, treatment can be stopped safely after the age of 3 years followed by further evaluation.

AUTOIMMUNE HYPOTHYROIDISM**Classification**

Autoimmune hypothyroidism may be associated with a goiter (Hashimoto's, or *goitrous thyroiditis*) or, at the later stages of the disease, minimal residual thyroid tissue (*atrophic thyroiditis*). Because the autoimmune process gradually reduces thyroid function, there is a phase of compensation when normal thyroid hormone levels are maintained by a rise in TSH. Although some patients may have minor symptoms, this state is called

subclinical hypothyroidism. Later, unbound T_4 levels fall and TSH levels rise further; symptoms become more readily apparent at this stage (usually TSH >10 mIU/L), which is referred to as *clinical hypothyroidism* or *overt hypothyroidism*.

Prevalence

The mean annual incidence rate of autoimmune hypothyroidism is up to 4 per 1000 women and 1 per 1000 men. It is more common in certain populations, such as the Japanese, probably because of genetic factors and chronic exposure to a high-iodine diet. The mean age at diagnosis is 60 years, and the prevalence of overt hypothyroidism increases with age. Subclinical hypothyroidism is found in 6–8% of women (10% over the age of 60) and 3% of men. The annual risk of developing clinical hypothyroidism is about 4% when subclinical hypothyroidism is associated with positive thyroid peroxidase (TPO) antibodies.

Pathogenesis

In Hashimoto's thyroiditis, there is a marked lymphocytic infiltration of the thyroid with germinal center formation, atrophy of the thyroid follicles accompanied by oxyphil metaplasia, absence of colloid, and mild to moderate fibrosis. In atrophic thyroiditis, the fibrosis is much more extensive, lymphocyte infiltration is less pronounced, and thyroid follicles are almost completely absent. Atrophic thyroiditis usually represents the end stage of Hashimoto's thyroiditis rather than a separate disorder, although a distinct form of marked fibrosis occurs in which the gland is infiltrated with IgG4-positive plasma cells.

As with most autoimmune disorders, susceptibility to autoimmune hypothyroidism is determined by a combination of genetic and environmental factors, and the risk of either autoimmune hypothyroidism or Graves' disease is increased among siblings. HLA-DR polymorphisms are the best documented genetic risk factors for autoimmune hypothyroidism, especially HLA-DR3, DR4, and DR5 in Caucasians. A weak association also exists between polymorphisms in *CTLA-4*, a T cell-regulatory gene, and autoimmune hypothyroidism. Both of these genetic associations are shared by other autoimmune diseases, which may explain the relationship between autoimmune hypothyroidism and other autoimmune diseases, especially type 1 diabetes mellitus, Addison's disease, pernicious anemia, and vitiligo. HLA-DR and *CTLA-4* polymorphisms account for approximately half of the genetic susceptibility to autoimmune hypothyroidism and the role of other contributory loci remains to be clarified. A gene on chromosome 21 may be responsible for the association between autoimmune hypothyroidism and Down's syndrome. The female preponderance of thyroid autoimmunity is most likely due to sex steroid effects on the immune response, but an X chromosome-related genetic factor is also possible and may account for the high frequency of autoimmune hypothyroidism in Turner's syndrome. Environmental susceptibility factors are poorly defined at present. A high iodine or low selenium intake and decreased exposure to microorganisms in childhood increase the risk of autoimmune hypothyroidism. Smoking cessation transiently increases incidence whereas alcohol intake seems protective. These factors may account for the increase in prevalence over the last two to three decades.

The thyroid lymphocytic infiltrate in autoimmune hypothyroidism is composed of activated T cells as well as B cells. Thyroid cell destruction is primarily mediated by the CD8⁺ cytotoxic T cells but local production of cytokines, such as tumor necrosis factor (TNF), interleukin-1 (IL-1), and interferon γ (IFN- γ), derived from the inflammatory infiltrate may render thyroid cells more susceptible to apoptosis mediated by death receptors, such as Fas, and by oxidative stress. These cytokines also impair thyroid cell function directly and induce the expression of other proinflammatory molecules by the thyroid cells themselves, such as cytokines, HLA class I and class II molecules, adhesion molecules, CD40, and nitric oxide. Administration of high concentrations of cytokines for therapeutic purposes (especially IFN- α) is associated with increased autoimmune thyroid disease, possibly through mechanisms similar to those in sporadic disease. Novel anticancer and immunomodulatory treatments, such as tyrosine kinase inhibitors and alemtuzumab, can also induce thyroid autoimmunity via their effects on T cell regulation.

Antibodies to TPO and thyroglobulin (Tg) are clinically useful markers of thyroid autoimmunity, but any pathogenic effect is restricted to a secondary role in amplifying an ongoing autoimmune response. TPO antibodies fix complement, and complement membrane-attack complexes are present in the thyroid in autoimmune hypothyroidism. However, transplacental passage of Tg or TPO antibodies has no effect on the fetal thyroid, which suggests that T cell-mediated injury is required to initiate autoimmune damage to the thyroid.

Up to 20% of patients with autoimmune hypothyroidism have antibodies against the TSH-R, which, in contrast to thyroid-stimulating immunoglobulin (TSI), do not stimulate the receptor but prevent the binding of TSH. These TSH-R-blocking antibodies, therefore, cause hypothyroidism and, especially in Asian patients, thyroid atrophy. Their transplacental passage may induce transient neonatal hypothyroidism. Rarely, patients have a mixture of TSI and TSH-R-blocking antibodies, and thyroid function can oscillate between hyperthyroidism and hypothyroidism as one or the other antibody becomes dominant. Predicting the course of disease in such individuals is difficult, and they require close monitoring of thyroid function. Bioassays can be used to document that TSH-R-blocking antibodies reduce the cyclic AMP-inducing effect of TSH on cultured TSH-R-expressing cells, but these assays are difficult to perform. Thyrotropin-binding inhibitory immunoglobulin (TBII) assays that measure the binding of antibodies to the receptor by competition with labeled TSH do not distinguish between TSI and TSH-R-blocking antibodies, but a positive result in a patient with spontaneous hypothyroidism is strong evidence for the presence of blocking antibodies. The use of these assays does not generally alter clinical management, although it may be useful to confirm the cause of transient neonatal hypothyroidism.

Clinical Manifestations

The main clinical features of hypothyroidism are summarized in [Table 376-3](#). The onset is usually insidious, and the patient may become aware of symptoms only when euthyroidism is restored. Patients with Hashimoto's thyroiditis may present because of goiter rather than symptoms of hypothyroidism. The goiter may not be large, but it is usually irregular and firm in consistency. Rarely uncomplicated Hashimoto's thyroiditis is associated with pain.

Patients with atrophic thyroiditis or the later stage of Hashimoto's thyroiditis present with symptoms and signs of hypothyroidism. The skin is dry, and there is decreased sweating, thinning of the epidermis, and hyperkeratosis of the stratum corneum. Increased dermal glycosaminoglycan content traps water, giving rise to skin thickening without pitting (*myxedema*). Typical features include a puffy face with edematous eyelids and nonpitting pretibial edema (**Fig. 376-1**). There is pallor, often with a yellow tinge to the skin due to carotene accumulation. Nail growth is retarded, and hair is dry, brittle, difficult to manage, and falls out easily. In addition to diffuse alopecia, there is thinning of the outer third of the eyebrows, although this is not a specific sign of hypothyroidism.

FIGURE 376-1

Facial appearance in hypothyroidism. Note puffy eyes and thickened skin.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Other common features include constipation and weight gain (despite a poor appetite). In contrast to popular perception, the weight gain is usually modest and due mainly to fluid retention in the myxedematous tissues. Libido is decreased in both sexes, and there may be oligomenorrhea or amenorrhea

in long-standing disease, but menorrhagia may occur at an early stage. Fertility is reduced, and the incidence of miscarriage is increased. Prolactin levels are often modestly increased (**Chap. 373**) and may contribute to alterations in libido and fertility and cause galactorrhea.

Myocardial contractility and pulse rate are reduced, leading to a reduced stroke volume and bradycardia. Increased peripheral resistance may be accompanied by hypertension, particularly diastolic. Blood flow is diverted from the skin, producing cool extremities. Pericardial effusions occur in up to 30% of patients but rarely compromise cardiac function. Although alterations in myosin heavy chain isoform expression have been documented, cardiomyopathy is rare. Fluid may also accumulate in other serous cavities and in the middle ear, giving rise to conductive deafness. Pulmonary function is generally normal, but dyspnea may be caused by pleural effusion, impaired respiratory muscle function, diminished ventilatory drive, or sleep apnea.

Carpal tunnel and other entrapment syndromes are common, as is impairment of muscle function with stiffness, cramps, and pain. On examination, there may be slow relaxation of tendon reflexes and pseudomyotonia. Memory and concentration are impaired. Experimentally, positron emission tomography (PET) scans examining glucose metabolism in hypothyroid subjects show lower regional activity in the amygdala, hippocampus, and perigenual anterior cingulate cortex, among other regions, and this activity corrects after thyroxine replacement. Rare neurologic problems include reversible cerebellar ataxia, dementia, psychosis, and myxedema coma. *Hashimoto's encephalopathy* has been defined as a steroid-responsive syndrome associated with TPO antibodies, myoclonus, and slow-wave activity on electroencephalography, but the relationship with thyroid autoimmunity or hypothyroidism is not established. The hoarse voice and occasionally clumsy speech of hypothyroidism reflect fluid accumulation in the vocal cords and tongue.

The features described above are the consequence of thyroid hormone deficiency. However, autoimmune hypothyroidism may be associated with signs or symptoms of other autoimmune diseases, particularly vitiligo, pernicious anemia, Addison's disease, alopecia areata, and type 1 diabetes mellitus. Less common associations include celiac disease, dermatitis herpetiformis, chronic active hepatitis, rheumatoid arthritis, systemic lupus erythematosus (SLE), myasthenia gravis, and Sjögren's syndrome. Thyroid-associated ophthalmopathy usually occurs in Graves' disease (see below), but in about 5% of patients it is associated with autoimmune hypothyroidism.

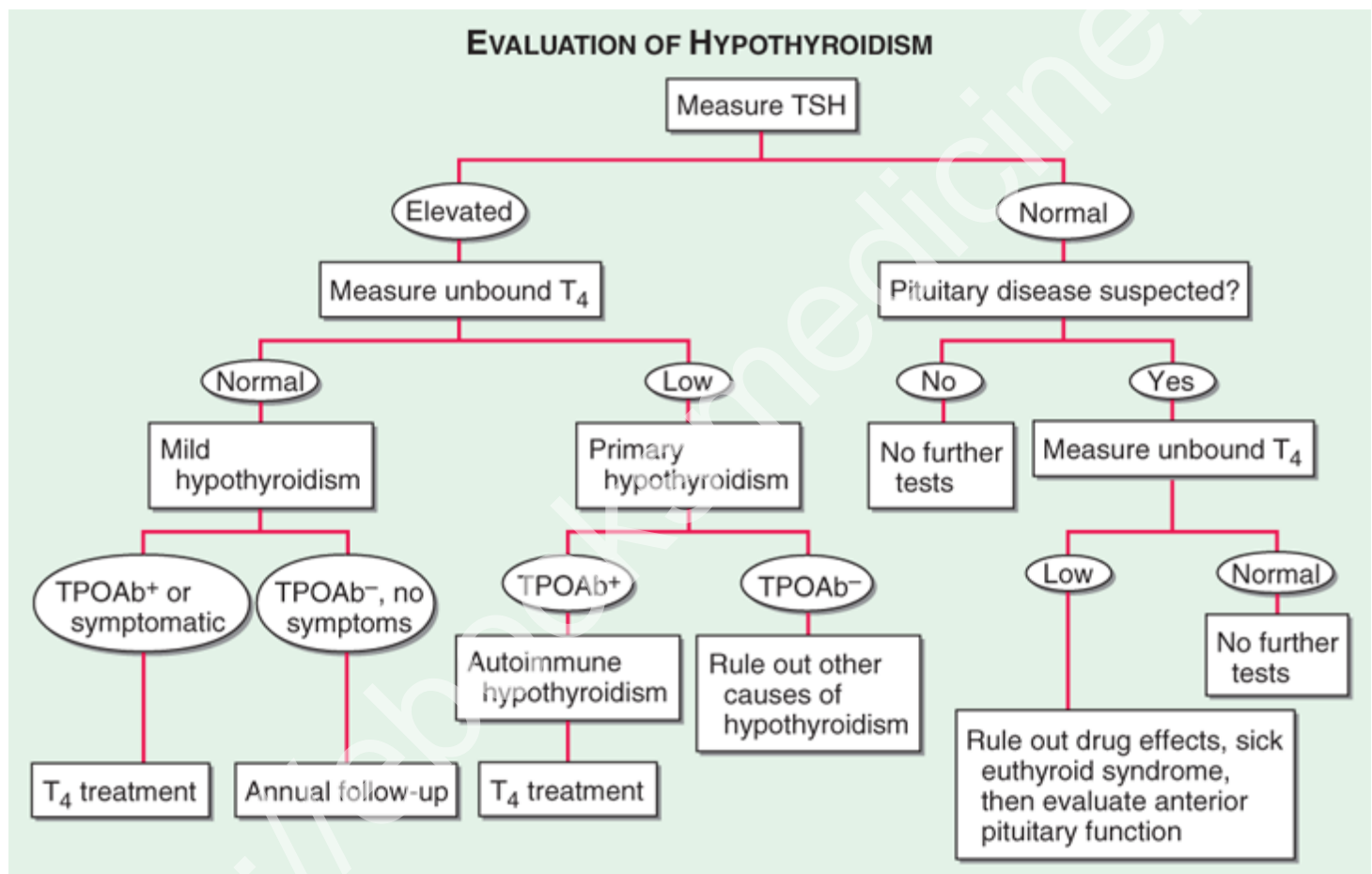
Autoimmune hypothyroidism is uncommon in children and usually presents with slow growth and delayed facial and dental maturation. The pituitary may be enlarged due to thyrotroph hyperplasia. Myopathy, with muscle swelling, is more common in children than in adults. In most cases, puberty is delayed, but precocious puberty sometimes occurs. There may be intellectual impairment if the onset is before 3 years and the hormone deficiency is severe.

Laboratory Evaluation

A summary of the investigations used to determine the existence and cause of hypothyroidism is provided in **Fig. 376-2**. A normal TSH level excludes primary (but not secondary) hypothyroidism. If the TSH is elevated, an unbound T_4 level is needed to confirm the presence of clinical hypothyroidism, but T_4 is inferior to TSH when used as a screening test, because it will not detect subclinical hypothyroidism. Circulating unbound T_3 levels are normal in about 25% of patients, reflecting adaptive deiodinase responses to hypothyroidism. T_3 measurements are, therefore, not indicated.

FIGURE 376-2

Evaluation of hypothyroidism. TPOAb⁺, thyroid peroxidase antibodies present; TPOAb⁻, thyroid peroxidase antibodies not present; TSH, thyroid-stimulating hormone.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.

Once clinical or subclinical hypothyroidism is confirmed, the etiology is usually easily established by demonstrating the presence of TPO and Tg antibodies, which are present in >95% of patients with autoimmune hypothyroidism. TBII can be found in 10–20% of patients, but measurement is not needed routinely. Other abnormal laboratory findings in hypothyroidism may include increased creatine phosphokinase, elevated cholesterol and triglycerides, and anemia (usually normocytic or macrocytic). Except when accompanied by iron deficiency, the anemia and other abnormalities gradually resolve with thyroxine replacement.

Differential Diagnosis

An asymmetric goiter in Hashimoto's thyroiditis may be confused with a multinodular goiter (MNG) or thyroid carcinoma, in which thyroid antibodies may also be present. Ultrasound can be used to show the presence of a solitary lesion or an MNG rather than the heterogeneous thyroid enlargement typical of Hashimoto's thyroiditis. FNA biopsy is useful in the investigation of focal nodules. Other causes of hypothyroidism are discussed below and in [Table 376-1](#) but rarely cause diagnostic confusion.

OTHER CAUSES OF HYPOTHYROIDISM

Iatrogenic hypothyroidism is a common cause of hypothyroidism and can often be detected by screening before symptoms develop. In the first 3–4 months after radioiodine treatment for Graves' disease, transient hypothyroidism may occur due to reversible radiation damage. Low-dose thyroxine treatment can be withdrawn if recovery occurs. Because TSH levels are suppressed by hyperthyroidism, unbound T₄ levels are a better measure of thyroid function than TSH in the months following radioiodine treatment. Mild hypothyroidism after subtotal thyroidectomy may also resolve after several months, as the gland remnant is stimulated by increased TSH levels.

Iodine deficiency is responsible for endemic goiter and cretinism but is an uncommon cause of adult hypothyroidism unless the iodine intake is very low or there are complicating factors, such as the consumption of thiocyanates in cassava or selenium deficiency. Although hypothyroidism due to iodine deficiency can be treated with thyroxine, public health measures to improve iodine intake should be advocated to eliminate this problem. Iodized salt or bread or a single bolus of oral or intramuscular iodized oil have all been used successfully.

Paradoxically, chronic iodine excess can also induce goiter and hypothyroidism. The intracellular events that account for this effect are unclear, but individuals with autoimmune thyroiditis are especially susceptible. Iodine excess is responsible for the hypothyroidism that occurs in up to 13% of patients treated with amiodarone (see below). Other drugs, particularly lithium, may also cause hypothyroidism. Transient hypothyroidism caused by thyroiditis is discussed below.

Secondary hypothyroidism is usually diagnosed in the context of other anterior pituitary hormone deficiencies; isolated TSH deficiency is very rare ([Chap. 372](#)). TSH levels may be low, normal, or even slightly increased in secondary hypothyroidism; the latter is due to secretion of immunoactive but bioinactive forms of TSH. The diagnosis is confirmed by detecting a low unbound T₄ level. The goal of treatment is to maintain T₄ levels in the upper half of the reference interval, because TSH levels cannot be used to monitor therapy.

TREATMENT

TREATMENT

Hypothyroidism

CLINICAL HYPOTHYROIDISM

If there is no residual thyroid function, the daily replacement dose of **levothyroxine** is usually 1.6 µg/kg body weight (typically 100–150 µg), ideally taken at least 30 min before breakfast. In many patients, however, lower doses suffice until residual thyroid tissue is destroyed. In patients who develop hypothyroidism after the treatment of Graves' disease, there is often underlying autonomous function, necessitating lower replacement doses (typically 75–125 µg/d).

Adult patients under 60 years old without evidence of heart disease may be started on 50–100 µg **levothyroxine** (T₄) daily. The dose is adjusted on the basis of TSH levels, with the goal of treatment being a normal TSH, ideally in the lower half of the reference range. TSH responses are gradual and should be measured about 2 months after instituting treatment or after any subsequent change in **levothyroxine** dosage. The clinical effects of **levothyroxine** replacement are slow to appear. Patients may not experience full relief from symptoms until 3–6 months after normal TSH levels are restored. Adjustment of **levothyroxine** dosage is made in 12.5- or 25-µg increments if the TSH is high; decrements of the same magnitude should be made if the TSH is suppressed. Patients with a suppressed TSH of any cause, including T₄ overtreatment, have an increased risk of atrial fibrillation and reduced bone density.

Although desiccated animal thyroid preparations (thyroid extract USP) are available, they are not recommended because the ratio of T₃ to T₄ is nonphysiologic. The use of **levothyroxine** combined with **liothyronine** (triiodothyronine, T₃) has been investigated, but benefit has not been confirmed in prospective studies. There is no place for **liothyronine** alone as long-term replacement, because the short half-life necessitates three or four daily doses and is associated with fluctuating T₃ levels.

Once full replacement is achieved and TSH levels are stable, follow-up measurement of TSH is recommended at annual intervals. It is important to ensure ongoing adherence as patients do not feel any symptomatic difference after missing a few doses of **levothyroxine**, and this sometimes leads to self-discontinuation.

In patients of normal body weight who are taking ≥200 µg of **levothyroxine** per day, an elevated TSH level is often a sign of poor adherence to treatment. This is also the likely explanation for fluctuating TSH levels, despite a constant **levothyroxine** dosage. Such patients often have normal or high unbound T₄ levels, despite an elevated TSH, because they remember to take medication for a few days before testing; this is sufficient to normalize T₄, but not TSH levels. It is important to consider variable adherence, because this pattern of thyroid function tests is otherwise suggestive of disorders associated with inappropriate TSH secretion (**Chap. 375**). Because T₄ has a long half-life (7 days), patients who miss a dose can be advised to take two doses of the skipped tablets at once. Other causes of increased **levothyroxine** requirements must be excluded, particularly malabsorption (e.g., celiac disease, small-bowel surgery, atrophic or *Helicobacter pylori*-related gastritis), oral estrogen containing medications or selective estrogen receptor modulator therapy, ingestion with a meal, and drugs that interfere with T₄ absorption or metabolism such as bile acid sequestrants, ferrous sulfate, calcium supplements, sevelamer, **sucralfate**, proton pump inhibitors,

lovastatin, aluminum hydroxide, rifampicin, amiodarone, carbamazepine, phenytoin, and tyrosine kinase inhibitors.

SUBCLINICAL HYPOTHYROIDISM

By definition, subclinical hypothyroidism refers to biochemical evidence of thyroid hormone deficiency in patients who have few or no apparent clinical features of hypothyroidism. There are no universally accepted recommendations for the management of subclinical hypothyroidism, but **levothyroxine** is recommended if the patient is a woman who wishes to conceive or is pregnant, or when TSH levels are above 10 mIU/L. Otherwise, when TSH levels are below 10 mIU/L, a trial of treatment may be considered when patients have suggestive symptoms of hypothyroidism, positive TPO antibodies, or any evidence of heart disease. It is important to confirm that any elevation of TSH is sustained over a 3-month period before treatment is given. Treatment is administered by starting with a low dose of **levothyroxine** (25–50 µg/d) with the goal of normalizing TSH. If **levothyroxine** is not given, thyroid function should be evaluated annually.

SPECIAL TREATMENT CONSIDERATIONS

Rarely, **levothyroxine** replacement is associated with pseudotumor cerebri in children. Presentation appears to be idiosyncratic and occurs months after treatment has begun.

Because maternal hypothyroidism may both adversely affect fetal neural development and be associated with adverse gestational outcomes (miscarriage, preterm delivery), thyroid function should be monitored to preserve euthyroidism in women with a history or high risk of hypothyroidism. The presence of thyroid autoantibodies alone, in a euthyroid patient, is also associated with miscarriage and preterm delivery; large-scale trials are underway to establish whether **levothyroxine** therapy improves outcomes in this group. Prior to conception, **levothyroxine** therapy should be targeted to maintain a serum TSH in the normal range but <2.5 mIU/L for hypothyroid women. Subsequently, thyroid function should be evaluated immediately after pregnancy is confirmed and every 4 weeks during the first half of the pregnancy, with less frequent testing after 20 weeks' gestation (every 6–8 weeks depending on whether **levothyroxine** dose adjustment is ongoing). The **levothyroxine** dose may need to be increased by up to 45% during pregnancy. Women should increase **levothyroxine** from once daily dosing to nine doses per week as soon as pregnancy is confirmed, to anticipate this change. Thereafter dosage should be closely monitored with a goal TSH in the lower half of the trimester-specific normative range, if available, or <2.5 mIU/L. After delivery, **levothyroxine** doses typically return to prepregnancy levels. Pregnant women should be counseled to separate ingestion of prenatal vitamins and iron supplements from **levothyroxine**.

Elderly patients may require 20% less thyroxine than younger patients. In the elderly, especially patients with known coronary artery disease, the starting dose of **levothyroxine** is 12.5–25 µg/d with similar increments every 2–3 months until TSH is normalized. In some patients, it may be impossible to achieve full replacement despite optimal antianginal treatment. *Emergency surgery* is generally safe in patients with untreated hypothyroidism, although routine surgery in a hypothyroid patient should be deferred until euthyroidism is achieved.

Myxedema coma still has a 20–40% mortality rate, despite intensive treatment, and outcomes are independent of the T_4 and TSH levels. Clinical manifestations include reduced level of consciousness, sometimes associated with seizures, as well as the other features of hypothyroidism ([Table 376-3](#)). Hypothermia can reach 23°C (74°F). There may be a history of treated hypothyroidism with poor compliance, or the patient may be previously undiagnosed. Myxedema coma almost always occurs in the elderly and is usually precipitated by factors that impair respiration, such as drugs (especially sedatives, anesthetics, and antidepressants), pneumonia, congestive heart failure, myocardial infarction, gastrointestinal bleeding, or cerebrovascular accidents. Sepsis should also be suspected. Exposure to cold may also be a risk factor. Hypoventilation, leading to hypoxia and hypercapnia, plays a major role in pathogenesis; hypoglycemia and dilutional hyponatremia also contribute to the development of myxedema coma.

[Levothyroxine](#) can initially be administered as a single IV bolus of 200–400 µg, which serves as a loading dose, followed by a daily oral dose of 1.6 µg/kg/d, reduced by 25% if administered IV. If suitable IV preparation is not available, the same initial dose of [levothyroxine](#) can be given by nasogastric tube (although absorption may be impaired in myxedema). Because $T_4 \rightarrow T_3$ conversion is impaired in myxedema coma, there is a rationale for adding [liothyronine](#) (T_3) intravenously or via nasogastric tube to [levothyroxine](#) treatment, although excess [liothyronine](#) has the potential to provoke arrhythmias. An initial loading dose of 5–20 µg [liothyronine](#) should be followed by 2.5–10 µg 8 hourly, with lower doses for those at cardiovascular risk.

Supportive therapy should be provided to correct any associated metabolic disturbances. External warming is indicated only if the temperature is <30°C, as it can result in cardiovascular collapse ([Chap. 454](#)). Space blankets should be used to prevent further heat loss. Parenteral [hydrocortisone](#) (50 mg every 6 h) should be administered, because there is impaired adrenal reserve in profound hypothyroidism. Any precipitating factors should be treated, including the early use of broad-spectrum antibiotics, pending the exclusion of infection. Ventilatory support with regular blood gas analysis is usually needed during the first 48 h. Hypertonic saline or IV glucose may be needed if there is severe hyponatremia or hypoglycemia; hypotonic IV fluids should be avoided because they may exacerbate water retention secondary to reduced renal perfusion and inappropriate vasopressin secretion. The metabolism of most medications is impaired, and sedatives should be avoided if possible or used in reduced doses. Medication blood levels should be monitored, when available, to guide dosage.

FURTHER READING

Hanley P et al: Thyroid disorders in children and adolescents: A review. *JAMA Pediatr* 170:1008, 2016. [[PubMed: 27571216](#)]

Jonklaas J et al: Guidelines for the treatment of hypothyroidism: Prepared by the American Thyroid Association Task Force on thyroid hormone replacement. *Thyroid* 24:1670, 2014. [[PubMed: 25266247](#)]

Rugge JB, Bougatsos C, Chou R. Screening and treatment of thyroid dysfunction: An evidence review for the U. S. Preventive Services Task Force. Ann Intern Med 162:35, 2015. [[PubMed: 25347444](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Harrison's Principles of Internal Medicine, 20e >

Chapter 377: Hyperthyroidism

J. Larry Jameson; Susan J. Mandel; Anthony P. Weetman

THYROTOXICOSIS

Thyrotoxicosis is defined as the state of thyroid hormone excess and is not synonymous with *hyperthyroidism*, which is the result of excessive thyroid function. However, the major etiologies of thyrotoxicosis are hyperthyroidism caused by Graves' disease, toxic multinodular goiter (MNG), and toxic adenomas. Other causes are listed in [Table 377-1](#).

TABLE 377-1

Causes of Thyrotoxicosis

Primary Hyperthyroidism
Graves' disease
Toxic multinodular goiter
Toxic adenoma
Functioning thyroid carcinoma metastases
Activating mutation of the TSH receptor
Activating mutation of $G_{S\alpha}$ (McCune-Albright syndrome)
Struma ovarii
Drugs: iodine excess (Jod-Basedow phenomenon)
Thyrotoxicosis without Hyperthyroidism
Subacute thyroiditis
Silent thyroiditis
Other causes of thyroid destruction: amiodarone, radiation, infarction of adenoma
Ingestion of excess thyroid hormone (thyrotoxicosis factitia) or thyroid tissue
Secondary Hyperthyroidism
TSH-secreting pituitary adenoma
Thyroid hormone resistance syndrome: occasional patients may have features of thyrotoxicosis
Chorionic gonadotropin-secreting tumors ^a
Gestational thyrotoxicosis ^a

^aCirculating TSH levels are low in these forms of secondary hyperthyroidism.

Abbreviation: TSH, thyroid-stimulating hormone.

GRAVES' DISEASE

Epidemiology

Graves' disease accounts for 60–80% of thyrotoxicosis. The prevalence varies among populations, reflecting genetic factors and iodine intake (high iodine intake is associated with an increased prevalence of Graves' disease). Graves' disease occurs in up to 2% of women but is one-tenth as frequent in men. The disorder rarely begins before adolescence and typically occurs between 20 and 50 years of age; it also occurs in the elderly.

Pathogenesis

As in autoimmune hypothyroidism, a combination of environmental and genetic factors, including polymorphisms in HLA-DR, the immunoregulatory genes *CTLA-4*, *CD25*, *PTPN22*, *FCRL3*, and *CD226*, as well as the gene encoding the thyroid-stimulating hormone receptor (TSH-R), contributes to Graves' disease susceptibility. The concordance for Graves' disease in monozygotic twins is 20–30%, compared to <5% in dizygotic twins. Indirect evidence suggests that stress is an important environmental factor, presumably operating through neuroendocrine effects on the immune system. Smoking is a minor risk factor for Graves' disease and a major risk factor for the development of ophthalmopathy. Sudden increases in iodine intake may precipitate Graves' disease, and there is a threefold increase in the occurrence of Graves' disease in the postpartum period. Graves' disease may occur during the immune reconstitution phase after highly active antiretroviral therapy (HAART) or alemtuzumab treatment.

The hyperthyroidism of Graves' disease is caused by thyroid-stimulating immunoglobulin (TSI) that are synthesized in the thyroid gland as well as in bone marrow and lymph nodes. Such antibodies can be detected by bioassays or by using the more widely available thyrotropin-binding inhibitory immunoglobulin (TBII) assays. The presence of TBII in a patient with thyrotoxicosis implies the existence of TSI, and these assays are useful in monitoring pregnant Graves' patients in whom high levels of TSI can cross the placenta and cause neonatal thyrotoxicosis. Other thyroid autoimmune responses, similar to those in autoimmune hypothyroidism (see above), occur concurrently in patients with Graves' disease. In particular, thyroid peroxidase (TPO) and thyroglobulin (Tg) antibodies occur in up to 80% of cases. Because the coexisting thyroiditis can also affect thyroid function, there is no direct correlation between the level of TSI and thyroid hormone levels in Graves' disease.

Cytokines appear to play a major role in thyroid-associated ophthalmopathy. There is infiltration of the extraocular muscles by activated T cells; the release of cytokines such as interferon γ (IFN- γ), tumor necrosis factor (TNF), and interleukin-1 (IL-1) results in fibroblast activation and increased synthesis of glycosaminoglycans that trap water, thereby leading to characteristic muscle swelling. Late in the disease,

there is irreversible fibrosis of the muscles. Though the pathogenesis of thyroid-associated ophthalmopathy remains unclear, there is mounting evidence that the TSH-R is a shared autoantigen that is expressed in the orbit; this would explain the close association with autoimmune thyroid disease. Increased fat is an additional cause of retrobulbar tissue expansion. The increase in intraorbital pressure can lead to proptosis, diplopia, and optic neuropathy.

Clinical Manifestations

Signs and symptoms include features that are common to any cause of thyrotoxicosis (**Table 377-2**) as well as those specific for Graves' disease. The clinical presentation depends on the severity of thyrotoxicosis, the duration of disease, individual susceptibility to excess thyroid hormone, and the patient's age. In the elderly, features of thyrotoxicosis may be subtle or masked, and patients may present mainly with fatigue and weight loss, a condition known as *apathetic thyrotoxicosis*.

TABLE 377-2

Signs and Symptoms of Thyrotoxicosis (Descending Order of Frequency)

SYMPTOMS	SIGNS ^a
Hyperactivity, irritability, dysphoria	Tachycardia; atrial fibrillation in the elderly
Heat intolerance and sweating	Tremor
Palpitations	Goiter
Fatigue and weakness	Warm, moist skin
Weight loss with increased appetite	Muscle weakness, proximal myopathy
Diarrhea	Lid retraction or lag
Polyuria	Gynecomastia
Oligomenorrhea, loss of libido	

^aExcludes the signs of ophthalmopathy and dermopathy specific for Graves' disease.

Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, due to the increased metabolic rate. Weight gain occurs in 5% of patients, however, because of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatigability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly. Fine tremor is a frequent finding, best elicited by having patients stretch out their fingers while feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is rare. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis, but it occurs in other ethnic groups as well.

The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations, occasionally caused by supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur and can lead to worsening of angina or heart failure in the elderly or those with preexisting heart disease. Atrial fibrillation is more common in patients >50 years of age. Treatment of the thyrotoxic state alone converts atrial fibrillation to normal sinus rhythm in about half of patients, suggesting the existence of an underlying cardiac problem in the remainder.

The skin is usually warm and moist, and the patient may complain of sweating and heat intolerance, particularly during warm weather. Palmar erythema, onycholysis, and, less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and a diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men, there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in long-standing thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalciuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

In Graves' disease, the thyroid is usually diffusely enlarged to two to three times its normal size. The consistency is firm, but not nodular. There may be a thrill or bruit, best detected at the inferolateral margins of the thyroid lobes, due to the increased vascularity of the gland and the hyperdynamic circulation.

Lid retraction, causing a staring appearance, can occur in any form of thyrotoxicosis and is the result of sympathetic overactivity. However, Graves' disease is associated with specific eye signs that comprise *Graves' ophthalmopathy* (**Fig. 377-1A**). This condition is also called *thyroid-associated ophthalmopathy*, because it occurs in the absence of hyperthyroidism in 10% of patients. Most of these individuals have autoimmune hypothyroidism or thyroid antibodies. The onset of Graves' ophthalmopathy occurs within the year before or after the diagnosis of thyrotoxicosis in 75% of patients but can sometimes precede or follow thyrotoxicosis by several years, accounting for some cases of euthyroid ophthalmopathy.

FIGURE 377-1

Features of Graves' disease. **A.** Ophthalmopathy in Graves' disease; lid retraction, periorbital edema, conjunctival injection, and proptosis are marked. **B.** Thyroid dermopathy over the lateral aspects of the shins. **C.** Thyroid acropachy.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Some patients with Graves' disease have little clinical evidence of ophthalmopathy. However, the enlarged extraocular muscles typical of the disease, and other subtle features, can be detected in most patients when investigated by ultrasound or computed tomography (CT) imaging of the orbits. Unilateral signs are found in up to 10% of patients. The earliest manifestations of ophthalmopathy are usually a sensation of grittiness, eye discomfort, and excess tearing. About one-third of patients have proptosis, best detected by visualization of the sclera between the lower border of the iris and the lower eyelid, with the eyes in the primary position. Proptosis can be measured using an exophthalmometer. In severe cases, proptosis may cause corneal exposure and damage, especially if the lids fail to close during sleep. Periorbital edema, scleral injection, and chemosis are also frequent. In 5–10% of patients, the muscle swelling is so severe that diplopia results, typically, but not exclusively, when the patient looks up and laterally. The most serious manifestation is compression of the optic nerve at the apex of the orbit, leading to papilledema; peripheral field defects; and, if left untreated, permanent loss of vision.

The “NO SPECS” scoring system to evaluate ophthalmopathy is an acronym derived from the following changes:

0 = No signs or symptoms

1 = **O**nly signs (lid retraction or lag), no symptoms

2 = **S**oft tissue involvement (periorbital edema)

3 = **P**roptosis (>22 mm)

4 = **E**xtraocular muscle involvement (diplopia)

5 = **C**orneal involvement

6 = **S**ight loss

Although useful as a mnemonic, the NO SPECS scheme is inadequate to describe the eye disease fully, and patients do not necessarily progress from one class to another; alternative scoring systems (e.g., the EUGOGO system developed by the European Group On Graves' Orbitopathy) that assess disease activity are preferable for monitoring and treatment purposes. When Graves' eye disease is active and severe, referral to an ophthalmologist is indicated and objective measurements are needed, such as lid-fissure width; corneal staining with fluorescein; and evaluation of extraocular muscle function (e.g., Hess chart), intraocular pressure and visual fields, acuity, and color vision.

Thyroid dermopathy occurs in <5% of patients with Graves' disease (**Fig. 377-1B**), almost always in the presence of moderate or severe ophthalmopathy. Although most frequent over the anterior and lateral aspects of the lower leg (hence the term *pretibial myxedema*), skin changes can occur at other sites, particularly after trauma. The typical lesion is a noninflamed, indurated plaque with a deep pink or purple color and an "orange skin" appearance. Nodular involvement can occur, and the condition can rarely extend over the whole lower leg and foot, mimicking elephantiasis. *Thyroid acropachy* refers to a form of clubbing found in <1% of patients with Graves' disease (**Fig. 377-1C**). It is so strongly associated with thyroid dermopathy that an alternative cause of clubbing should be sought in a Graves' patient without coincident skin and orbital involvement.

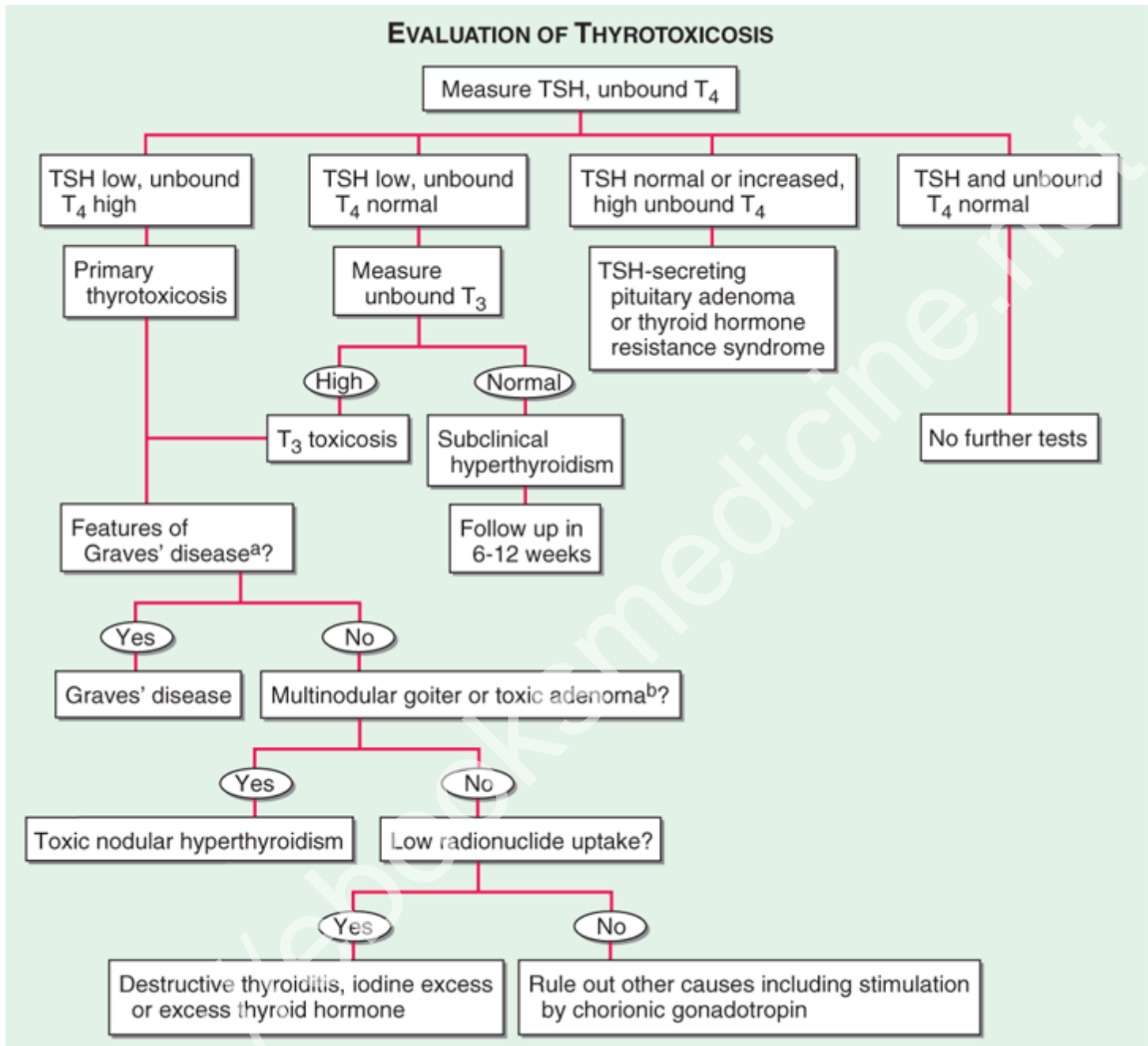
Laboratory Evaluation

Investigations used to determine the existence and cause of thyrotoxicosis are summarized in **Fig. 377-2**. In Graves' disease, the TSH level is suppressed, and total and unbound thyroid hormone levels are increased. In 2–5% of patients (and more in areas of borderline iodine intake), only T₃ is increased (T₃ toxicosis). The converse state of T₄ toxicosis, with elevated total and unbound T₄ and normal T₃ levels, is occasionally seen when hyperthyroidism is induced by excess iodine, providing surplus substrate for thyroid hormone synthesis. Measurement of TPO antibodies or TBII may be useful if the diagnosis is unclear clinically but is not needed routinely. Associated abnormalities that may cause diagnostic confusion in thyrotoxicosis include elevation of bilirubin, liver enzymes, and ferritin. Microcytic anemia and thrombocytopenia may occur.

FIGURE 377-2

Evaluation of thyrotoxicosis. ^aDiffuse goiter, positive TPO antibodies or TRAb, ophthalmopathy, dermopathy.

^bCan be confirmed by radionuclide scan. TSH, thyroid-stimulating hormone.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Differential Diagnosis

Diagnosis of Graves' disease is straightforward in a patient with biochemically confirmed thyrotoxicosis, diffuse goiter on palpation, ophthalmopathy, and often a personal or family history of autoimmune disorders. For patients with thyrotoxicosis who lack these features, the diagnosis is generally established by a radionuclide (^{99m}Tc , ^{123}I , or ^{131}I) scan and uptake of the thyroid, which will distinguish the diffuse, high uptake of Graves' disease from destructive thyroiditis, ectopic thyroid tissue, and factitious thyrotoxicosis, as

well as diagnosing a toxic adenoma or toxic MNG. Alternatively, TRAb measurement can be used to diagnose Graves' disease and color-flow Doppler ultrasonography may distinguish between hyperthyroidism (with increased blood flow) and destructive thyroiditis. In secondary hyperthyroidism due to a TSH-secreting pituitary tumor, there is also a diffuse goiter. The presence of a nonsuppressed TSH level and the finding of a pituitary tumor on CT or magnetic resonance imaging (MRI) scan suggest this diagnosis.

Clinical features of thyrotoxicosis can mimic certain aspects of other disorders, including panic attacks, mania, pheochromocytoma, and weight loss associated with malignancy. The diagnosis of thyrotoxicosis can be easily excluded if the TSH and unbound T_4 and T_3 levels are normal. A normal TSH also excludes Graves' disease as a cause of diffuse goiter.

Clinical Course

Clinical features generally worsen without treatment; mortality was 10–30% before the introduction of satisfactory therapy. Some patients with mild Graves' disease experience spontaneous relapses and remissions. Rarely, there may be fluctuation between hypo- and hyperthyroidism due to changes in the functional activity of TSH-R antibodies. About 15% of patients who enter remission after treatment develop hypothyroidism 10–15 years later as a result of the destructive autoimmune process.

The clinical course of ophthalmopathy does not follow that of the thyroid disease, although thyroid dysfunction can worsen eye signs. Ophthalmopathy typically worsens over the initial 3–6 months, followed by a plateau phase over the next 12–18 months, and then some spontaneous improvement, particularly in the soft tissue changes. However, the course is more fulminant in up to 5% of patients, requiring intervention in the acute phase if there is optic nerve compression or corneal ulceration. Diplopia may appear late in the disease due to fibrosis of the extraocular muscles. Radioiodine treatment for hyperthyroidism worsens the eye disease in a small proportion of patients (especially smokers). Antithyroid drugs or surgery have no adverse effects on the clinical course of ophthalmopathy. Thyroid dermopathy, when it occurs, usually appears 1–2 years after the development of Graves' hyperthyroidism; it may improve spontaneously.

TREATMENT

TREATMENT

Graves' Disease

The *hyperthyroidism* of Graves' disease is treated by reducing thyroid hormone synthesis, using an antithyroid drug, or reducing the amount of thyroid tissue with radioiodine (^{131}I) treatment or by thyroidectomy. Antithyroid drugs are the predominant therapy in many centers in Europe, Latin America, and Japan, whereas radioiodine is more often the first line of treatment in North America. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve remission.

The main *antithyroid drugs* are thionamides; propylthiouracil, carbimazole (not available in the United States), and the active metabolite of the latter, methimazole. All inhibit the function of TPO, reducing oxidation and organification of iodide. These drugs also reduce thyroid antibody levels by mechanisms that remain unclear, and they appear to enhance spontaneous rates of remission. Propylthiouracil inhibits deiodination of $T_4 \rightarrow T_3$. However, this effect is of minor benefit, except in the most severe thyrotoxicosis, and is offset by the much shorter half-life of this drug (90 min) compared to methimazole (6 h). Due to the hepatotoxicity of propylthiouracil, the U.S. Food and Drug Administration (FDA) has limited indications for its use to the first trimester of pregnancy, the treatment of thyroid storm, and patients with minor adverse reactions to methimazole. If propylthiouracil is used, monitoring of liver function tests is recommended.

There are many variations of antithyroid drug regimens. The initial dose of carbimazole or methimazole is usually 10–20 mg every 8 or 12 h, but once-daily dosing is possible after euthyroidism is restored. Propylthiouracil is given at a dose of 100–200 mg every 6–8 h, and divided doses are usually given throughout the course. Lower doses of each drug may suffice in areas of low iodine intake. The starting dose of an antithyroid drug can be gradually reduced (titration regimen) as thyrotoxicosis improves. Less commonly, high doses may be given combined with *levothyroxine* supplementation (block-replace regimen) to avoid drug-induced hypothyroidism. The titration regimen is preferred to minimize the dose of antithyroid drug and provide an index of treatment response.

Thyroid function tests and clinical manifestations are reviewed 4–6 weeks after starting treatment, and the dose is titrated based on unbound T_4 levels. Most patients do not achieve euthyroidism until 6–8 weeks after treatment is initiated. TSH levels often remain suppressed for several months and therefore do not provide a sensitive index of treatment response. The usual daily maintenance doses of antithyroid drugs in the titration regimen are 2.5–10 mg of carbimazole or methimazole and 50–100 mg of propylthiouracil. In the block-replace regimen, the initial dose of antithyroid drug is held constant, and the dose of *levothyroxine* is adjusted to maintain normal unbound T_4 levels. When TSH suppression is alleviated, TSH levels can also be used to monitor therapy.

Maximum remission rates (up to 30–60% in some populations) are achieved by 12–18 months for the titration regimen and are higher in patients where TRAb levels are no longer detected, than in those with TRAb persistence. For unclear reasons, remission rates appear to vary in different geographic regions. Younger patients, males, smokers, and patients with a history of allergy, severe hyperthyroidism or large goiters are most likely to relapse when treatment stops, but outcomes are difficult to predict. All patients should be followed closely for relapse during the first year after treatment and at least annually thereafter.

The common minor side effects of antithyroid drugs are rash, urticaria, fever, and arthralgia (1–5% of patients). These may resolve spontaneously or after substituting an alternative antithyroid drug; rashes may respond to an antihistamine. Rare but major side effects include hepatitis (especially with propylthiouracil; avoid use in children) and cholestasis (methimazole and carbimazole); vasculitis; and, most important, agranulocytosis (<1%). It is essential that antithyroid drugs are stopped and not restarted if a patient develops major side effects. Written instructions should be provided regarding the symptoms of possible

agranulocytosis (e.g., sore throat, fever, mouth ulcers) and the need to stop treatment pending an urgent complete blood count to confirm that agranulocytosis is not present. Management of agranulocytosis is described in [Chap. 98](#). It is not useful to monitor blood counts prospectively, because the onset of agranulocytosis is idiosyncratic and abrupt.

Propranolol (20–40 mg every 6 h) or longer-acting selective β_1 receptor blockers such as atenolol may be helpful to control adrenergic symptoms, especially in the early stages before antithyroid drugs take effect. Beta blockers are also useful in patients with thyrotoxic periodic paralysis, pending correction of thyrotoxicosis. In consultation with a cardiologist, anticoagulation with warfarin should be considered in all patients with atrial fibrillation; there is often spontaneous reversion to sinus rhythm with control of hyperthyroidism, and long-term anticoagulation is not usually needed. Decreased warfarin doses are required when patients are thyrotoxic. If digoxin is used, increased doses are often needed in the thyrotoxic state.

Radioiodine causes progressive destruction of thyroid cells and can be used as initial treatment or for relapses after a trial of antithyroid drugs. There is a small risk of thyrotoxic crisis (see below) after radioiodine, which can be minimized by pretreatment with antithyroid drugs for at least a month before treatment. Antecedent treatment with an antithyroid drug and a beta blocker should be considered for all elderly patients or for those with cardiac problems. Carbimazole or methimazole must be stopped 2–3 days before radioiodine administration to achieve optimum iodine uptake, and can be restarted 3–7 days after radioiodine in those at risk of complications from worsening thyrotoxicosis. Propylthiouracil appears to have a prolonged radioprotective effect and should be stopped for a longer period before radioiodine is given, or a larger dose of radioiodine will be necessary.

Efforts to calculate an optimal dose of radioiodine that achieves euthyroidism without a high incidence of relapse or progression to hypothyroidism have not been successful. Some patients inevitably relapse after a single dose because the biologic effects of radiation vary between individuals, and hypothyroidism cannot be uniformly avoided even using accurate dosimetry. A practical strategy is to give a fixed dose based on clinical features, such as the severity of thyrotoxicosis, the size of the goiter (increases the dose needed), and the level of radioiodine uptake (decreases the dose needed). ^{131}I dosage generally ranges between 370 MBq (10 mCi) and 555 MBq (15 mCi). Most authorities favor an approach aimed at thyroid ablation (as opposed to euthyroidism), given that [levothyroxine](#) replacement is straightforward and most patients ultimately progress to hypothyroidism over 5–10 years, frequently with some delay in the diagnosis of hypothyroidism.

Certain radiation safety precautions are necessary in the first few days after radioiodine treatment, but the exact guidelines vary depending on local protocols. In general, patients need to avoid close, prolonged contact with children and pregnant women for 5–7 days because of possible transmission of residual isotope and exposure to radiation emanating from the gland. Rarely, there may be mild pain due to radiation thyroiditis 1–2 weeks after treatment. Hyperthyroidism can persist for 2–3 months before radioiodine takes full effect. For this reason, β -adrenergic blockers or antithyroid drugs can be used to control symptoms during this interval. Persistent hyperthyroidism can be treated with a second dose of radioiodine, usually 6

months after the first dose. The risk of hypothyroidism after radioiodine depends on the dosage but is at least 10–20% in the first year and 5% per year thereafter. Patients should be informed of this possibility before treatment and require close follow-up during the first year followed by annual thyroid function testing.

Pregnancy and breast-feeding are absolute contraindications to radioiodine treatment, but patients can conceive safely 6 months after treatment. The presence of ophthalmopathy, especially in smokers, requires caution. [Prednisone](#), 30 mg/d, at the time of radioiodine treatment, tapered over 6–8 weeks may prevent exacerbation of ophthalmopathy, but radioiodine should generally be avoided in those with active moderate to severe eye disease. The overall risk of cancer after radioiodine treatment in adults is not increased. Although many physicians avoid radioiodine in children and adolescents because of the theoretical risks of malignancy, emerging evidence suggests that radioiodine can be used safely in older children.

Total or near-total thyroidectomy is an option for patients who relapse after antithyroid drugs and prefer this treatment to radioiodine. Some experts recommend surgery in young individuals, particularly when the goiter is very large. Careful control of thyrotoxicosis with antithyroid drugs, followed by potassium iodide (1–2 drops SSKI orally tid for 10 days), is needed prior to surgery to avoid thyrotoxic crisis and to reduce the vascularity of the gland. The major complications of surgery—bleeding, laryngeal edema, hypoparathyroidism, and damage to the recurrent laryngeal nerves—are unusual when the procedure is performed by highly experienced surgeons. Recurrence rates in the best series are <2%, but the rate of hypothyroidism is similar to that following radioiodine treatment, especially with the current trend away from subtotal thyroidectomy.

Antithyroid drugs should be used to manage Graves' disease in *pregnancy*. Because transplacental passage of these drugs may produce fetal hypothyroidism and goiter if the maternal dose is excessive, maternal antithyroid dose titration should target serum free or total T₄ levels at or just above the pregnancy reference range. If available, propylthiouracil should be used until 14–16 weeks' gestation because of the association of rare cases of methimazole/carbimazole embryopathy, including *aplasia cutis* and other defects, such as choanal atresia and tracheoesophageal fistulae. Because of the potential for teratogenic effects, recent recommendations suggest discontinuation of antithyroid medication in a newly pregnant woman with Graves' disease, who is euthyroid on a low dose of methimazole (<5–10 mg/day) or PTU (<100–200 mg/day), after evaluating recent thyroid function tests, disease history, goiter size, duration of therapy, and TRAb measurement. Following cessation, careful monitoring of maternal thyroid function tests is essential. On the other hand, for women at high risk of developing thyrotoxicosis if antithyroid drugs are discontinued (large goiter, requirement for higher antithyroid drug dosage), continued therapy is necessary, with PTU (if available) administration in the first trimester. But, because of its rare association with hepatotoxicity, propylthiouracil should be limited to the first trimester and then maternal therapy should be converted to methimazole (or carbimazole) at a ratio of 15–20 mg of propylthiouracil to 1 mg of methimazole. It is often possible to stop treatment in the last trimester because TSIs tend to decline in pregnancy. Nonetheless, the transplacental transfer of these antibodies if present at levels 3 times higher than the normative range rarely causes *fetal or neonatal thyrotoxicosis*. Poor intrauterine growth, a fetal heart rate of >160 beats/min,

advanced bone age, fetal goiter, and high levels of maternal TSI after 26 weeks gestation may herald this complication. Antithyroid drugs given to the mother can be used to treat the fetus and may be needed for 1–3 months after delivery, until the maternal antibodies disappear from the baby's circulation. The postpartum period is a time of major risk for relapse of Graves' disease. Breast-feeding is safe with low doses of antithyroid drugs. Graves' disease in *children* is usually managed initially with methimazole or carbimazole (avoid propylthiouracil), often given as a prolonged course of the titration regimen. Surgery or radioiodine may be indicated for severe or relapsing disease.

Thyrotoxic crisis, or *thyroid storm*, is rare and presents as a life-threatening exacerbation of hyperthyroidism, accompanied by fever, delirium, seizures, coma, vomiting, diarrhea, and jaundice. The mortality rate due to cardiac failure, arrhythmia, or hyperthermia is as high as 30%, even with treatment. Thyrotoxic crisis is usually precipitated by acute illness (e.g., stroke, infection, trauma, diabetic ketoacidosis), surgery (especially on the thyroid), or radioiodine treatment of a patient with partially treated or untreated hyperthyroidism. Management requires intensive monitoring and supportive care, identification and treatment of the precipitating cause, and measures that reduce thyroid hormone synthesis. Large doses of propylthiouracil (500–1000 mg loading dose and 250 mg every 4 h) should be given orally or by nasogastric tube or per rectum; the drug's inhibitory action on $T_4 \rightarrow T_3$ conversion makes it the antithyroid drug of choice. If not available, methimazole can be used in doses of 20 mg every 6 h. One hour after the first dose of propylthiouracil, stable iodide (5 drops SSKI every 6 h) is given to block thyroid hormone synthesis via the Wolff-Chaikoff effect (the delay allows the antithyroid drug to prevent the excess iodine from being incorporated into new hormone). Propranolol should also be given to reduce tachycardia and other adrenergic manifestations (60–80 mg PO every 4 h; or 2 mg IV every 4 h). Although other β -adrenergic blockers can be used, high doses of propranolol decrease $T_4 \rightarrow T_3$ conversion, and the doses can be easily adjusted. Caution is needed to avoid acute negative inotropic effects, but controlling the heart rate is important, as some patients develop a form of high-output heart failure. Short-acting IV esmolol can be used to decrease heart rate while monitoring for signs of heart failure. Additional therapeutic measures include glucocorticoids (e.g., [hydrocortisone](#) 300 mg IV bolus, then 100 mg every 8 h), antibiotics if infection is present, cholestyramine to sequester thyroid hormones, cooling, oxygen, and IV fluids.

Ophthalmopathy requires no active treatment when it is mild or moderate, because there is usually spontaneous improvement. General measures include meticulous control of thyroid hormone levels, cessation of smoking, and an explanation of the natural history of ophthalmopathy. Discomfort can be relieved with artificial tears (e.g., hypromellose 0.3% or carbomer 0.2% ophthalmic gel) paraffin-based eye ointment, and the use of dark glasses with side frames. Periorbital edema may respond to a more upright sleeping position or a diuretic. Corneal exposure during sleep can be avoided by using patches or taping the eyelids shut. Minor degrees of diplopia improve with prisms fitted to spectacles. Some authorities also advocate selenium 100 μ g bd. Severe ophthalmopathy, with optic nerve involvement or chemosis resulting in corneal damage, is an emergency requiring joint management with an ophthalmologist. Pulse therapy with IV methylprednisolone (e.g., 500 mg of methylprednisolone once weekly for 6 weeks, then 250 mg once weekly for 6 weeks) is preferable to oral glucocorticoids, which are used for moderately active disease. When glucocorticoids are ineffective, orbital decompression can be achieved by removing bone from any wall of

the orbit, thereby allowing displacement of fat and swollen extraocular muscles. The transantral route is used most often because it requires no external incision. Proptosis recedes an average of 5 mm, but there may be residual or even worsened diplopia. Once the eye disease has stabilized, surgery may be indicated for relief of diplopia and correction of the appearance. External beam radiotherapy of the orbits has been used for many years, but the efficacy of this therapy remains unclear, and it is best reserved for those with moderately active disease who have failed or are not candidates for glucocorticoid therapy. Other immunosuppressive agents such as rituximab have shown some benefit, but their role is yet to be established.

Thyroid dermopathy does not usually require treatment, but it can cause cosmetic problems or interfere with the fit of shoes. Surgical removal is not indicated. If necessary, treatment consists of topical, high-potency glucocorticoid ointment under an occlusive dressing. Octreotide may be beneficial in some cases.

OTHER CAUSES OF THYROTOXICOSIS

Destructive thyroiditis (subacute or silent thyroiditis) typically presents with a short thyrotoxic phase due to the release of preformed thyroid hormones and catabolism of Tg (see “[Subacute Thyroiditis](#),” below). True hyperthyroidism is absent, as demonstrated by a low radionuclide uptake. Circulating Tg levels are typically increased. Other causes of thyrotoxicosis with low or absent thyroid radionuclide uptake include *thyrotoxicosis factitia*, iodine excess, and, rarely, ectopic thyroid tissue, particularly teratomas of the ovary (*struma ovarii*) and functional metastatic follicular carcinoma. Whole-body radionuclide studies can demonstrate ectopic thyroid tissue, and thyrotoxicosis factitia can be distinguished from destructive thyroiditis by the clinical features and low levels of Tg. Amiodarone treatment is associated with thyrotoxicosis in up to 10% of patients, particularly in areas of low iodine intake (see below).

TSH-secreting pituitary adenoma is a rare cause of thyrotoxicosis. It is characterized by the presence of an inappropriately normal or increased TSH level in a patient with hyperthyroidism, diffuse goiter, and elevated T₄ and T₃ levels ([Chap. 373](#)). Elevated levels of the α -subunit of TSH, released by the TSH-secreting adenoma, support this diagnosis, which can be confirmed by demonstrating the pituitary tumor on MRI or CT scan. A combination of transsphenoidal surgery, sella irradiation, and octreotide may be required to normalize TSH, because many of these tumors are large and locally invasive at the time of diagnosis. Radioiodine or antithyroid drugs can be used to control thyrotoxicosis.

Thyrotoxicosis caused by *toxic MNG* and *hyperfunctioning solitary nodules* is discussed below.

THYROIDITIS

A clinically useful classification of thyroiditis is based on the onset and duration of disease ([Table 377-3](#)).

TABLE 377-3

Causes of Thyroiditis

Acute
Bacterial infection: especially <i>Staphylococcus</i> , <i>Streptococcus</i> , and <i>Enterobacter</i>
Fungal infection: <i>Aspergillus</i> , <i>Candida</i> , <i>Coccidioides</i> , <i>Histoplasma</i> , and <i>Pneumocystis</i>
Radiation thyroiditis after ^{131}I treatment
Amiodarone (may also be subacute or chronic)
Subacute
Viral (or granulomatous) thyroiditis
Silent thyroiditis (including postpartum thyroiditis)
Mycobacterial infection
Drug induced (interferon, amiodarone)
Chronic
Autoimmunity: focal thyroiditis, Hashimoto's thyroiditis, atrophic thyroiditis
Riedel's thyroiditis
Parasitic thyroiditis: echinococcosis, strongyloidiasis, cysticercosis
Traumatic: after palpation

ACUTE THYROIDITIS

Acute thyroiditis is rare and due to suppurative infection of the thyroid. In children and young adults, the most common cause is the presence of a piriform sinus, a remnant of the fourth branchial pouch that connects the oropharynx with the thyroid. Such sinuses are predominantly left-sided. A long-standing goiter and degeneration in a thyroid malignancy are risk factors in the elderly. The patient presents with thyroid pain, often referred to the throat or ears, and a small, tender goiter that may be asymmetric. Fever,

dysphagia, and erythema over the thyroid are common, as are systemic symptoms of a febrile illness and lymphadenopathy.

The differential diagnosis of *thyroid pain* includes subacute or, rarely, chronic thyroiditis; hemorrhage into a cyst; malignancy including lymphoma; and, rarely, amiodarone-induced thyroiditis or amyloidosis. However, the abrupt presentation and clinical features of acute thyroiditis rarely cause confusion. The erythrocyte sedimentation rate (ESR) and white cell count are usually increased, but thyroid function is normal. Fine-needle aspiration (FNA) biopsy shows infiltration by polymorphonuclear leukocytes; culture of the sample can identify the organism. Caution is needed in immunocompromised patients as fungal, mycobacterial, or *Pneumocystis* thyroiditis can occur in this setting. Antibiotic treatment is guided initially by Gram stain and, subsequently, by cultures of the FNA biopsy. Surgery may be needed to drain an abscess, which can be localized by CT scan or ultrasound. Tracheal obstruction, septicemia, retropharyngeal abscess, mediastinitis, and jugular venous thrombosis may complicate acute thyroiditis but are uncommon with prompt use of antibiotics.

SUBACUTE THYROIDITIS

This is also termed *de Quervain's thyroiditis*, *granulomatous thyroiditis*, or *viral thyroiditis*. Many viruses have been implicated, including mumps, coxsackie, influenza, adenoviruses, and echoviruses, but attempts to identify the virus in an individual patient are often unsuccessful and do not influence management. The diagnosis of subacute thyroiditis is often overlooked because the symptoms can mimic pharyngitis. The peak incidence occurs at 30–50 years, and women are affected three times more frequently than men.

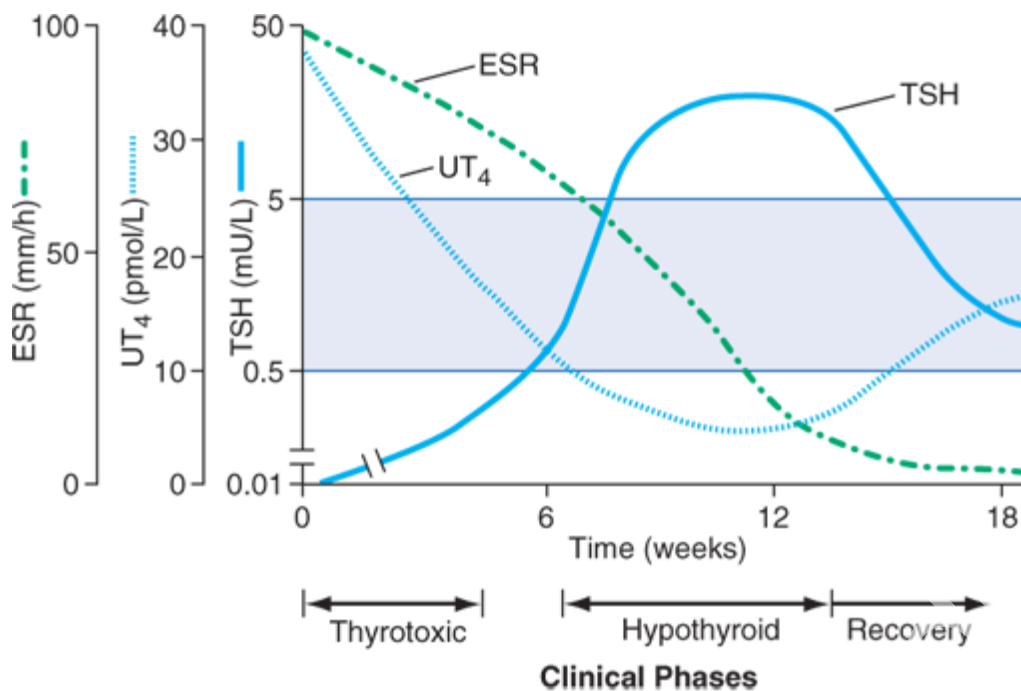
Pathophysiology

The thyroid shows a characteristic patchy inflammatory infiltrate with disruption of the thyroid follicles and multinucleated giant cells within some follicles. The follicular changes progress to granulomas accompanied by fibrosis. Finally, the thyroid returns to normal, usually several months after onset. During the initial phase of follicular destruction, there is release of Tg and thyroid hormones, leading to increased circulating T₄ and T₃ and suppression of TSH (Fig. 377-3). During this destructive phase, radioactive iodine uptake is low or undetectable. After several weeks, the thyroid is depleted of stored thyroid hormone and a phase of hypothyroidism typically occurs, with low unbound T₄ (and sometimes T₃) and moderately increased TSH levels. Radioactive iodine uptake returns to normal or is even increased as a result of the rise in TSH. Finally, thyroid hormone and TSH levels return to normal as the disease subsides.

FIGURE 377-3

Clinical course of subacute thyroiditis. The release of thyroid hormones is initially associated with a thyrotoxic phase and suppressed thyroid-stimulating hormone (TSH). A hypothyroid phase then ensues, with low T₄ and TSH levels that are initially low but gradually increase. During the recovery phase, increased TSH levels combined with resolution of thyroid follicular injury lead to normalization of thyroid function, often

several months after the beginning of the illness. ESR, erythrocyte sedimentation rate; UT_4 , free or unbound T_4 .



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Clinical Manifestations

The patient usually presents with a painful and enlarged thyroid, sometimes accompanied by fever. There may be features of thyrotoxicosis or hypothyroidism, depending on the phase of the illness. Malaise and symptoms of an upper respiratory tract infection may precede the thyroid-related features by several weeks. In other patients, the onset is acute, severe, and without obvious antecedent. The patient typically complains of a sore throat, and examination reveals a small goiter that is exquisitely tender. Pain is often referred to the jaw or ear. Complete resolution is the usual outcome, but late-onset permanent hypothyroidism occurs in 15% of cases, particularly in those with coincidental thyroid autoimmunity. A prolonged course over many months, with one or more relapses, occurs in a small percentage of patients.

Laboratory Evaluation

As depicted in Fig. 377-3, thyroid function tests characteristically evolve through three distinct phases over about 6 months: (1) thyrotoxic phase, (2) hypothyroid phase, and (3) recovery phase. In the thyrotoxic phase, T_4 and T_3 levels are increased, reflecting their discharge from the damaged thyroid cells, and TSH is suppressed. The T_4/T_3 ratio is greater than in Graves' disease or thyroid autonomy, in which T_3 is often disproportionately increased. The diagnosis is confirmed by a high ESR and low uptake of radioiodine (<5%) or ^{99m}Tc pertechnetate (as compared to salivary gland pertechnetate concentration). The white blood cell

count may be increased, and thyroid antibodies are negative. If the diagnosis is in doubt, FNA biopsy may be useful, particularly to distinguish unilateral involvement from bleeding into a cyst or neoplasm.

TREATMENT

TREATMENT

Subacute Thyroiditis

Relatively large doses of aspirin (e.g., 600 mg every 4–6 h) or nonsteroidal anti-inflammatory drugs (NSAIDs) are sufficient to control symptoms in many cases. If this treatment is inadequate, or if the patient has marked local or systemic symptoms, glucocorticoids should be given. The usual starting dose is 15–40 mg of [prednisone](#), depending on severity. The dose is gradually tapered over 6–8 weeks, in response to improvement in symptoms and the ESR. If a relapse occurs during glucocorticoid withdrawal, the dosage should be increased and then withdrawn more gradually. Thyroid function should be monitored every 2–4 weeks using TSH and unbound T₄ levels. Symptoms of thyrotoxicosis improve spontaneously but may be ameliorated by β -adrenergic blockers; antithyroid drugs play no role in treatment of the thyrotoxic phase. [Levothyroxine](#) replacement may be needed if the hypothyroid phase is prolonged, but doses should be low enough (50–100 μ g daily) to allow TSH-mediated recovery.

SILENT THYROIDITIS

Painless thyroiditis, or “*silent*” *thyroiditis*, occurs in patients with underlying autoimmune thyroid disease and has a clinical course similar to that of subacute thyroiditis. The condition occurs in up to 5% of women 3–6 months after pregnancy and is then termed *postpartum thyroiditis*. Typically, patients have a brief phase of thyrotoxicosis lasting 2–4 weeks, followed by hypothyroidism for 4–12 weeks, and then resolution; often, however, only one phase is apparent. The condition is associated with the presence of TPO antibodies antepartum, and it is three times more common in women with type 1 diabetes mellitus. As in subacute thyroiditis, the uptake of ^{99m}Tc pertechnetate or radioactive iodine is initially suppressed. In addition to the painless goiter, silent thyroiditis can be distinguished from subacute thyroiditis by a normal ESR and the presence of TPO antibodies. Glucocorticoid treatment is not indicated for silent thyroiditis. Severe thyrotoxic symptoms can be managed with a brief course of propranolol, 20–40 mg three or four times daily. Thyroxine replacement may be needed for the hypothyroid phase but should be withdrawn after 6–9 months, as recovery is the rule. Annual follow-up thereafter is recommended, because a proportion of these individuals develop permanent hypothyroidism. The condition may recur in subsequent pregnancies.

DRUG-INDUCED THYROIDITIS

Patients receiving cytokines, such as IFN- α or IL-2, or tyrosine kinase inhibitors may develop painless thyroiditis. IFN- α , which is used to treat chronic hepatitis B or C and hematologic and skin malignancies, causes thyroid dysfunction in up to 5% of treated patients. It has been associated with painless thyroiditis,

hypothyroidism, and Graves' disease, and is most common in women with TPO antibodies prior to treatment. For discussion of amiodarone, see "[Amiodarone Effects on Thyroid Function](#)," below.

CHRONIC THYROIDITIS

Focal thyroiditis is present in 20–40% of euthyroid autopsy cases and is associated with serologic evidence of autoimmunity, particularly the presence of TPO antibodies. The most common clinically apparent cause of chronic thyroiditis is *Hashimoto's thyroiditis*, an autoimmune disorder that often presents as a firm or hard goiter of variable size (see above). *Riedel's thyroiditis* is a rare disorder that typically occurs in middle-aged women. It presents with an insidious, painless goiter with local symptoms due to compression of the esophagus, trachea, neck veins, or recurrent laryngeal nerves. Dense fibrosis disrupts normal gland architecture and can extend outside the thyroid capsule. Despite these extensive histologic changes, thyroid dysfunction is uncommon. The goiter is hard, nontender, often asymmetric, and fixed, leading to suspicion of a malignancy. Diagnosis requires open biopsy as FNA biopsy is usually inadequate. Treatment is directed to surgical relief of compressive symptoms. Tamoxifen may also be beneficial. There is an association between Riedel's thyroiditis and IgG4-related disease causing idiopathic fibrosis at other sites (retroperitoneum, mediastinum, biliary tree, lung, and orbit).

SICK EUTHYROID SYNDROME (NONTHYROIDAL ILLNESS)

Any acute, severe illness can cause abnormalities of circulating TSH or thyroid hormone levels in the absence of underlying thyroid disease, making these measurements potentially misleading. The major cause of these hormonal changes is the release of cytokines such as IL-6. Unless a thyroid disorder is strongly suspected, the routine testing of thyroid function should be avoided in acutely ill patients.

The most common hormone pattern in sick euthyroid syndrome (SES), also called nonthyroidal illness (NTI), is a decrease in total and unbound T_3 levels (low T_3 syndrome) with normal levels of T_4 and TSH. The magnitude of the fall in T_3 correlates with the severity of the illness. T_4 conversion to T_3 via peripheral 5' (outer ring) deiodination is impaired, leading to increased reverse T_3 (rT_3). Since rT_3 is metabolized by 5' deiodination, its clearance is also reduced. Thus, decreased clearance rather than increased production is the major basis for increased rT_3 . Also, T_4 is alternately metabolized to the hormonally inactive T_3 sulfate. It is generally assumed that this low T_3 state is adaptive, because it can be induced in normal individuals by fasting. Teleologically, the fall in T_3 may limit catabolism in starved or ill patients.

Very sick patients may exhibit a dramatic fall in total T_4 and T_3 levels (low T_4 syndrome). With decreased tissue perfusion, muscle and liver expression of the type 3 deiodinase leads to accelerated T_4 and T_3 metabolism. This state has a poor prognosis. Another key factor in the fall in T_4 levels is altered binding to thyroxine-binding globulin (TBG). The commonly used free T_4 assays are subject to artifact when serum binding proteins are low and underestimate the true free T_4 level. Fluctuation in TSH levels also creates

challenges in the interpretation of thyroid function in sick patients. TSH levels may range from <0.1 mIU/L in very ill patients, especially with dopamine or glucocorticoid therapy, to >20 mIU/L during the recovery phase of SES. The exact mechanisms underlying the subnormal TSH seen in 10% of sick patients and the increased TSH seen in 5% remain unclear but may be mediated by cytokines including IL-12 and IL-18.

Any severe illness can induce changes in thyroid hormone levels, but certain disorders exhibit a distinctive pattern of abnormalities. Acute liver disease is associated with an initial rise in total (but not unbound) T_3 and T_4 levels due to TBG release; these levels become subnormal with progression to liver failure. A transient increase in total and unbound T_4 levels, usually with a normal T_3 level, is seen in 5–30% of acutely ill psychiatric patients. TSH values may be transiently low, normal, or high in these patients. In the early stage of HIV infection, T_3 and T_4 levels rise, even if there is weight loss. T_3 levels fall with progression to AIDS, but TSH usually remains normal. Renal disease is often accompanied by low T_3 concentrations, but with normal rather than increased rT_3 levels, due to an unknown factor that increases uptake of rT_3 into the liver.

The diagnosis of SES is challenging. Historic information may be limited, and patients often have multiple metabolic derangements. Useful features to consider include previous history of thyroid disease and thyroid function tests, evaluation of the severity and time course of the patient's acute illness, documentation of medications that may affect thyroid function or thyroid hormone levels, and measurements of rT_3 together with unbound thyroid hormones and TSH. The diagnosis of SES is frequently presumptive, given the clinical context and pattern of laboratory values; only resolution of the test results with clinical recovery can clearly establish this disorder. Treatment of SES with thyroid hormone (T_4 and/or T_3) is controversial, but most authorities recommend monitoring the patient's thyroid function tests during recovery, without administering thyroid hormone, unless there is historic or clinical evidence suggestive of hypothyroidism. Sufficiently large randomized controlled trials using thyroid hormone are unlikely to resolve this therapeutic controversy in the near future, because clinical presentations and outcomes are highly variable.

AMIODARONE EFFECTS ON THYROID FUNCTION

Amiodarone is a commonly used type III antiarrhythmic agent ([Chap. 247](#)). It is structurally related to thyroid hormone and contains 39% iodine by weight. Thus, typical doses of amiodarone (200 mg/d) are associated with very high iodine intake, leading to greater than fortyfold increases in plasma and urinary iodine levels. Moreover, because amiodarone is stored in adipose tissue, high iodine levels persist for >6 months after discontinuation of the drug. Amiodarone inhibits deiodinase activity, and its metabolites function as weak antagonists of thyroid hormone action. Amiodarone has the following effects on thyroid function: (1) acute, transient suppression of thyroid function; (2) hypothyroidism in patients susceptible to the inhibitory effects of a high iodine load; and (3) thyrotoxicosis that may be caused by either a Jod-Basedow effect from the iodine load, in the setting of MNG or incipient Graves' disease, or a thyroiditis-like condition.

The initiation of amiodarone treatment is associated with a transient decrease of T_4 levels, reflecting the inhibitory effect of iodine on T_4 release. Soon thereafter, most individuals escape from iodide-dependent suppression of the thyroid (Wolff-Chaikoff effect), and the inhibitory effects on deiodinase activity and thyroid hormone receptor action become predominant. These events lead to the following pattern of thyroid function tests: increased T_4 , decreased T_3 , increased rT_3 , and a transient TSH increase (up to 20 mIU/L). TSH levels normalize or are slightly suppressed within 1–3 months.

The incidence of hypothyroidism from amiodarone varies geographically, apparently correlating with iodine intake. Hypothyroidism occurs in up to 13% of amiodarone-treated patients in iodine-replete countries, such as the United States, but is less common (<6% incidence) in areas of lower iodine intake, such as Italy or Spain. The pathogenesis appears to involve an inability of the thyroid gland to escape from the Wolff-Chaikoff effect in autoimmune thyroiditis. Consequently, amiodarone-associated hypothyroidism is more common in women and individuals with positive TPO antibodies. It is usually unnecessary to discontinue amiodarone for this side effect, because [levothyroxine](#) can be used to normalize thyroid function. TSH levels should be monitored, because T_4 levels are often increased for the reasons described above.

The management of amiodarone-induced thyrotoxicosis (AIT) is complicated by the fact that there are different causes of thyrotoxicosis and because the increased thyroid hormone levels exacerbate underlying arrhythmias and coronary artery disease. Amiodarone treatment causes thyrotoxicosis in 10% of patients living in areas of low iodine intake and in 2% of patients in regions of high iodine intake. There are two major forms of AIT, although some patients have features of both. Type 1 AIT is associated with an underlying thyroid abnormality (preclinical Graves' disease or nodular goiter). Thyroid hormone synthesis becomes excessive as a result of increased iodine exposure (Jod-Basedow phenomenon). Type 2 AIT occurs in individuals with no intrinsic thyroid abnormalities and is the result of drug-induced lysosomal activation leading to destructive thyroiditis with histiocyte accumulation in the thyroid; the incidence rises as cumulative amiodarone dosage increases. Mild forms of type 2 AIT can resolve spontaneously or can occasionally lead to hypothyroidism. Color-flow Doppler ultrasonography shows increased vascularity in type 1 AIT but decreased vascularity in type 2 AIT. Thyroid scintiscans are difficult to interpret in this setting because the high endogenous iodine levels diminish tracer uptake. However, the presence of normal or rarely increased uptake favors type 1 AIT.

In AIT, the drug should be stopped, if possible, although this is often impractical because of the underlying cardiac disorder. Discontinuation of amiodarone will not have an acute effect because of its storage and prolonged half-life. High doses of antithyroid drugs can be used in type 1 AIT but are often ineffective. Potassium perchlorate, 200 mg every 6 h, has been used to reduce thyroidal iodide content. Perchlorate treatment has been associated with agranulocytosis, although the risk appears relatively low with short-term use. Glucocorticoids, as administered for subacute thyroiditis, have modest benefit in type 2 AIT. Lithium blocks thyroid hormone release and can also provide some benefit. Near-total thyroidectomy rapidly decreases thyroid hormone levels and may be the most effective long-term solution if the patient can undergo the procedure safely.

FURTHER READING

Biondi B et al: The 2015 European Thyroid Association Guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. Eur Thyroid J 4:149, 2015. [[PubMed: 26558232](#)]

De Leo S et al: Hyperthyroidism. Lancet 388:906, 2016. [[PubMed: 27038492](#)]

Hanley P et al: Thyroid disorders in children and adolescents: A review. JAMA Pediatr 170:1008, 2016. [[PubMed: 27571216](#)]

Ross DS et al: 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. Thyroid 26:1343, 2016.

Wiersinga WM: Advances in treatment of active, moderate-to-severe Graves' ophthalmopathy. Lancet Diabetes Endocrinol 5:134, 2017.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 449: Heavy Metal Poisoning

Howard Hu

INTRODUCTION

Metals pose a significant threat to health through low-level environmental as well as occupational exposures. One indication of their importance relative to other potential hazards is their ranking by the U.S. Agency for Toxic Substances and Disease Registry, which maintains an updated list of all hazards present in toxic waste sites according to their prevalence and the severity of their toxicity. The first, second, third, and seventh hazards on the list are heavy metals: lead, mercury, arsenic, and cadmium, respectively (<http://www.atsdr.cdc.gov/spl/>). Specific information pertaining to each of these metals, including sources and metabolism, toxic effects produced, diagnosis, and the appropriate treatment for poisoning, is summarized in **Table 449-1**.

TABLE 449-1

Heavy Metals

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
Arsenic				
Smelting and microelectronics industries; wood preservatives, pesticides, herbicides, fungicides; contaminant of deep-water wells; folk remedies; and coal; incineration of these products.	Organic arsenic (arsenobetaine, arsenocholine) is ingested in seafood and fish, but is nontoxic; inorganic arsenic is readily absorbed (lung and GI); sequesters in liver, spleen, kidneys, lungs, and GI tract; residues persist in skin, hair, and nails; biomethylation results in detoxification, but this process saturates.	Acute arsenic poisoning results in necrosis of intestinal mucosa with hemorrhagic gastroenteritis, fluid loss, hypotension, delayed cardiomyopathy, acute tubular necrosis, and hemolysis. Chronic arsenic exposure causes diabetes, vasospasm, peripheral vascular insufficiency and gangrene, peripheral neuropathy, and cancer of skin, lung, liver (angiosarcoma), bladder, and kidney. Lethal dose: 120–200 mg (adults); 2 mg/kg (children).	Nausea, vomiting, diarrhea, abdominal pain, delirium, coma, seizures; garlicky odor on breath; hyperkeratosis, hyperpigmentation, exfoliative dermatitis, and Mees' lines (transverse white striae of the fingernails); sensory and motor polyneuritis, distal weakness. Radiopaque sign on abdominal x-ray; ECG–QRS broadening, QT prolongation, ST depression, T-wave flattening; 24-h urinary arsenic >67 $\mu\text{mol/d}$ or 50 $\mu\text{g/d}$; (no seafood \times 24 h); if recent exposure, serum arsenic >0.9 $\mu\text{mol/L}$ (7 $\mu\text{g/dL}$). High arsenic in hair or nails.	If acute ingestion, ipecac to induce vomiting, gastric lavage, activated charcoal with a cathartic. Supportive care in ICU. Dimercaprol 3–5 mg/kg IM q4h \times 2 days; q6h \times 1 day, then q12h \times 10 days; alternative: oral succimer .

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
Cadmium				
Metal-plating, pigment, smelting, battery, and plastics industries; tobacco; incineration of these products; ingestion of food that concentrates cadmium (grains, cereals).	Absorbed through ingestion or inhalation; bound by metallothionein, filtered at the glomerulus, but reabsorbed by proximal tubules (thus, poorly excreted). Biologic half-life: 10–30 y. Binds cellular sulfhydryl groups, competes with zinc, calcium for binding sites. Concentrates in liver and kidneys.	Acute cadmium inhalation causes pneumonitis after 4–24 h; acute ingestion causes gastroenteritis. Chronic exposure causes anosmia, yellowing of teeth, emphysema, minor LFT elevations, microcytic hypochromic anemia unresponsive to iron therapy, proteinuria, increased urinary β_2 -microglobulin, calciuria, leading to chronic renal failure, osteomalacia, and fractures. Possible risks of cardiovascular disease and cancer.	With inhalation: pleuritic chest pain, dyspnea, cyanosis, fever, tachycardia, nausea, noncardiogenic pulmonary edema. With ingestion: nausea, vomiting, cramps, diarrhea. Bone pain, fractures with osteomalacia. If recent exposure, serum cadmium >500 nmol/L (5 $\mu\text{g/dL}$). Urinary cadmium >100 nmol/L (10 $\mu\text{g/g}$ creatinine) and/or urinary β_2 -microglobulin >750 $\mu\text{g/g}$ creatinine (but urinary β_2 -microglobulin also increased in other renal diseases such as pyelonephritis).	There is no effective treatment for cadmium poisoning (chelation not useful; dimercaprol can exacerbate nephrotoxicity). Avoidance of further exposure, supportive therapy, vitamin D for osteomalacia.
Lead				
Manufacturing of auto batteries, lead crystal,	Absorbed through ingestion or inhalation;	Acute exposure with blood lead levels (BPb) of >60–80 $\mu\text{g/dL}$ can	Abdominal pain, irritability, lethargy, anorexia, anemia, Fanconi's syndrome,	Identification and correction of exposure sources is critical. In some U.S. states, screening and

ceramics, fishing MAIN SOURCES weights, etc.;	organic lead METABOLISM (e.g., tetraethyl	TOXICITY	DIAGNOSIS	reporting to local health TREATMENT boards of children with
demolition or sanding of lead-painted houses, bridges; stained glass-making, plumbing, soldering; environmental exposure to paint chips, house dust (in homes built <1975), firing ranges (from bullet dust), food or water from improperly glazed ceramics, lead pipes; contaminated herbal remedies, candies; exposure to the combustion of leaded fuels.	lead) absorbed dermally. In blood, 95–99% sequestered in RBCs—thus, must measure lead in whole blood (not serum). Distributed widely in soft tissue, with half-life ~30 days; 15% of dose sequestered in bone with half-life of >20 years. Excreted mostly in urine, but also appears in other fluids including breast milk. Interferes with mitochondrial oxidative phosphorylation, ATPases, calcium-dependent messengers; enhances oxidation and cell apoptosis.	cause impaired neurotransmission and neuronal cell death (with central and peripheral nervous system effects); impaired hematopoiesis and renal tubular dysfunction. At higher levels of exposure (e.g., BPb >80–120 µg/dL), acute encephalopathy with convulsions, coma, and death may occur. Subclinical exposures in children (BPb 25–60 µg/dL) are associated with anemia; mental retardation; and deficits in language, motor function, balance, hearing, behavior, and school performance. Impairment of IQ appears to occur at even lower levels of exposure with no measurable threshold above	pyuria, azotemia in children with blood lead level (BPb) >80 µg/dL; may also see epiphyseal plate “lead lines” on long bone x-rays. Convulsions, coma at BPb >120 µg/dL. Noticeable neurodevelopmental delays at BPb of 40–80 µg/dL; may also see symptoms associated with higher BPb levels. Screening of all U.S. children when they begin to crawl (~6 months) is recommended by the CDC; source identification and intervention is begun if the BPb >10 µg/dL. In adults, acute exposure causes similar symptoms as in children as well as headaches, arthralgias, myalgias, depression, impaired short-term memory, loss of libido. Physical examination may	BPb >10 µg/dL and workers with BPb >40 µg/dL is required. In the highly exposed individual with symptoms, chelation is recommended with oral DMSA (succimer); if acutely toxic, hospitalization and IV or IM chelation with ethylenediaminetetraacetic acid calcium disodium (CaEDTA) may be required, with the addition of dimercaprol to prevent worsening of encephalopathy. It is uncertain whether children with asymptomatic lead exposure (e.g., BPb 20–40 µg/dL) benefit from chelation; a recent randomized trial showed no benefit. Correction of dietary deficiencies in iron, calcium, magnesium, and zinc will lower lead absorption and may also improve toxicity. Vitamin C is a weak but natural chelating agent. Calcium supplements (1200 mg at bedtime) have been shown to lower blood lead levels in pregnant women.

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
		<p>the limit of detection in most assays of 1 µg/dL. In adults, chronic subclinical exposures (BPb >40 µg/dL) are associated with an increased risk of anemia, demyelinating peripheral neuropathy (mainly motor), impairments of reaction time and hearing, accelerated declines in cognition, hypertension, ECG conduction delays, higher risk of cardiovascular disease and death, interstitial nephritis and chronic renal failure, diminished sperm counts, and spontaneous abortions.</p>	<p>reveal a “lead line” at the gingiva-tooth border, pallor, wrist drop, and cognitive dysfunction (e.g., declines on the mini-mental state exam); lab tests may reveal a normocytic, normochromic anemia, basophilic stippling, an elevated blood protoporphyrin level (free erythrocyte or zinc), and motor delays on nerve conduction. U.S. OSHA requires regular testing of lead-exposed workers with removal if BPb >40 µg/dL. New guidelines have been proposed recommending that BPb be maintained at <10 µg/dL, removal of workers if BPb >20 µg/dL, and monitoring of cumulative exposure parameters.</p>	
Mercury				
Metallic, mercurous, and	Elemental mercury (Hg) is	Acute inhalation of Hg vapor causes	Chronic exposure to metallic mercury	Treat acute ingestion of mercuric salts with induced

mercuric MAIN SOURCES mercury (Hg,	METABOLISM	TOXICITY	DIAGNOSIS	emesis or gastric lavage TREATMENT and polythiol resins (to
<p>Hg⁺, Hg²⁺)</p> <p>exposures occur in some chemical, metal-processing, electrical-equipment, automotive industries; they are also in thermometers, dental amalgams, batteries. Mercury is dispersed by waste incineration. Environmental bacteria convert inorganic to organic mercury, which then bioconcentrates up the aquatic food chain to contaminate tuna, swordfish, and other pelagic fish.</p>	<p>not well absorbed; however, it will volatilize into highly absorbable vapor. Inorganic mercury is absorbed through the gut and skin. Organic mercury is well absorbed through inhalation and ingestion. Elemental and organic mercury cross the blood-brain barrier and placenta. Mercury is excreted in urine and feces and has a half-life in blood of ~60 days; however, deposits will remain in the kidney and brain for years. Exposure to mercury stimulates the kidney to produce metallothionein, which provides</p>	<p>pneumonitis and noncardiogenic pulmonary edema leading to death, CNS symptoms, and polyneuropathy. Chronic high exposure causes CNS toxicity (mercurial <i>erethism</i>; see Diagnosis); lower exposures impair renal function, motor speed, memory, coordination. Acute ingestion of inorganic mercury causes gastroenteritis, the nephritic syndrome, or acute renal failure, hypertension, tachycardia, and cardiovascular collapse, with death at a dose of 10–42 mg/kg. Ingestion of organic mercury causes gastroenteritis, arrhythmias, and lesions in the basal ganglia, gray</p>	<p>vapor produces a characteristic intention tremor and mercurial <i>erethism</i>: excitability, memory loss, insomnia, timidity, and delirium (“mad as a hatter”). On neurobehavioral tests: decreased motor speed, visual scanning, verbal and visual memory, visuomotor coordination. Children exposed to mercury in any form may develop <i>acrodynia</i> (“pink disease”): flushing, itching, swelling, tachycardia, hypertension, excessive salivation or perspiration, irritability, weakness, morbilliform rashes, desquamation of palms and soles. Toxicity from elemental or inorganic mercury exposure begins when blood levels >180 nmol/L (3.6 µg/dL) and urine</p>	<p>bind mercury in the GI tract). Chelate with dimercaprol (up to 24 mg/kg per day IM in divided doses), DMSA (succimer), or penicillamine, with 5-day courses separated by several days of rest. If renal failure occurs, treat with peritoneal dialysis, hemodialysis, or extracorporeal regional complexing hemodialysis and succimer. Chronic inorganic mercury poisoning is best treated with <i>N</i>-acetyl penicillamine.</p>

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
	some detoxification benefit. Mercury binds sulfhydryl groups and interferes with a wide variety of critical enzymatic processes.	matter, and cerebellum at doses >1.7 mg/kg. High exposure during pregnancy causes derangement of fetal neuronal migration resulting in severe mental retardation. Mild exposures during pregnancy (from fish consumption) are associated with declines in neurobehavioral performance in offspring. Dimethylmercury, a compound only found in research labs, is “supertoxic”—a few drops of exposure via skin absorption or inhaled vapor can cause severe cerebellar degeneration and death.	levels >0.7 $\mu\text{mol/L}$ (15 $\mu\text{g/dL}$). Exposures that ended years ago may result in a >20- μg increase in 24-h urine after a 2-g dose of succimer . Organic mercury exposure is best measured by levels in blood (if recent) or hair (if chronic); CNS toxicity in children may derive from fetal exposures associated with maternal hair Hg >30 nmol/g (6 $\mu\text{g/g}$).	

Abbreviations: ATPase, adenosine triphosphatase; BPb, blood lead; CDC, Centers for Disease Control and Prevention; CNS, central nervous system; DMSA, dimercaptosuccinic acid; ECG, electrocardiogram; GI, gastrointestinal; ICU,

intensive care unit; IQ, intelligence quotient; LFT, liver function tests; OSHA, Occupational Safety and Health Administration; RBC, red blood cell.

Metals are inhaled primarily as dusts and fumes (the latter defined as tiny particles generated by combustion). Metal poisoning can also result from exposure to vapors (e.g., mercury vapor in creating dental amalgams). When metals are ingested in contaminated food or drink or by hand-to-mouth activity (implicated especially in children), their gastrointestinal absorption varies greatly with the specific chemical form of the metal and the nutritional status of the host. Once a metal is absorbed, blood is the main medium for its transport, with the precise kinetics dependent on diffusibility, protein binding, rates of biotransformation, availability of intracellular ligands, and other factors. Some organs (e.g., bone, liver, and kidney) sequester metals in relatively high concentrations for years. Most metals are excreted through renal clearance and gastrointestinal excretion; some proportion is also excreted through salivation, perspiration, exhalation, lactation, skin exfoliation, and loss of hair and nails. The intrinsic stability of metals facilitates tracing and measurement in biologic material, although the clinical significance of the levels measured is not always clear.

Some metals, such as [copper](#) and selenium, are essential to normal metabolic function as trace elements ([Chap. 326](#)) but are toxic at high levels of exposure. Others, such as lead and mercury, are xenobiotic and theoretically are capable of exerting toxic effects at any level of exposure. Indeed, much research is currently focused on the contribution of low-level xenobiotic metal exposure to chronic diseases and to subtle changes in health that may have significant public health consequences. Genetic factors, such as polymorphisms that encode for variant enzymes with altered properties in terms of metal binding, transport, and effects, also may modify the impact of metals on health and thereby account, at least in part, for individual susceptibility to metal effects.

The most important component of treatment for metal toxicity is the termination of exposure. *Chelating agents* are used to bind metals into stable cyclic compounds with relatively low toxicity and to enhance their excretion. The principal chelating agents are [dimercaprol](#) (British anti-Lewisite [BAL]), ethylenediamine tetraacetic acid (EDTA), [succimer](#) (dimercaptosuccinic acid [DMSA]), and penicillamine; their specific use depends on the metal involved and the clinical circumstances. Activated charcoal does not bind metals and thus is of limited usefulness in cases of acute metal ingestion.

In addition to the information provided in [Table 449-1](#), several other aspects of exposure, toxicity, or management are worthy of discussion with respect to the four most hazardous toxicants (arsenic, cadmium, lead, and mercury).

Arsenic, even at moderate levels of exposure, has been clearly linked with increased risks for cancer of the skin, bladder, renal pelvis, ureter, kidney, liver, and lung. These risks appear to be modified by smoking, folate and selenium status, genetic traits (such as ability to methylate arsenic), and other factors. Recent studies in community-based populations have generated strong evidence that arsenic exposure is a risk factor for increased coronary heart disease and stroke, lung function impairment, acute respiratory tract


infections, respiratory symptoms, and non-malignant lung disease mortality. Evidence is also emerging that low-level arsenic may cause neurodevelopmental delays in children and diabetes.

Serious *cadmium* poisoning from the contamination of food and water by mining effluents in Japan contributed to the 1946 outbreak of “itai-itai” (“ouch-ouch”) disease, so named because of cadmium-induced bone toxicity that led to painful bone fractures. Modest exposures from environmental contamination have been associated in some studies with a lower bone density, a higher incidence of fractures, and a faster decline in height in both men and women, effects that may be related to cadmium’s calciuric effect on the kidney. There is some evidence for synergy between the adverse impacts of cadmium and lead on kidney function. Environmental exposures have also been linked to lower lung function (even after adjusting for smoking cigarettes, which contain cadmium) as well as increased risk of cardiovascular disease and mortality, stroke, and heart failure. Several studies have also raised concerns that cadmium may be carcinogenic and contribute to elevated risks of prostate, breast, and pancreatic cancer. Overall, this growing body of research indicates that cadmium exposure may be contributing significantly to morbidity and mortality rates in the general population.

Advances in our understanding of *lead* toxicity have recently benefited by the development of K x-ray fluorescence (KXRF) instruments for making safe in vivo measurements of lead levels in bone (which, in turn, reflect cumulative exposure over many years, as opposed to blood lead levels, which mostly reflect recent exposure). Higher bone lead levels measured by KXRF have been linked to increased risk of hypertension and accelerated declines in cognition in both men and women from an urban population. Upon reviewing these studies in conjunction with other epidemiologic and toxicologic studies, a recent federal expert panel concluded that the impact of lead exposure on hypertension and cognition in adults was causal. Prospective studies have also demonstrated that higher bone lead levels are a major risk factor for increased cardiovascular morbidity and mortality rates in both community-based and occupational-exposed populations. Lead exposure at community levels has also been recently associated with increased risks of hearing loss, Parkinson’s disease, and amyotrophic lateral sclerosis. With respect to pregnancy-associated risks, high maternal bone lead levels were found to predict lower birth weight, head circumference, birth length, and neurodevelopmental performance in offspring by age 2 years. Offspring have also been shown to have higher blood pressures at age 7–14 years, an age range at which higher blood pressures are known to predict an elevated risk of developing hypertension. In a randomized trial, calcium supplementation (1200 mg daily) was found to significantly reduce the mobilization of lead from maternal bone into blood during pregnancy.

The toxicity of low-level organic *mercury* exposure (as manifested by neurobehavioral performance) is of increasing concern based on studies of the offspring of mothers who ingested mercury-contaminated fish. With respect to whether the consumption of fish by women during pregnancy is good or bad for offspring neurodevelopment, balancing the trade-offs of the beneficial effects of the omega-3-fatty acids (FAs) in fish versus the adverse effects of mercury contamination in fish has led to some confusion and inconsistency in public health recommendations. Overall, it would appear that it would be best for pregnant women to either limit fish consumption to those species known to be low in mercury contamination but high in omega-3-FAs

(such as sardines or mackerel) or to avoid fish and obtain omega-3-FAs through supplements or other dietary sources. Accumulated evidence has not supported the contention that ethyl mercury, used as a preservative in multiuse vaccines administered in early childhood, has played a significant role in causing neurodevelopmental problems such as autism. With regard to adults, there is conflicting evidence as to whether mercury exposure is associated with increased risk of hypertension and cardiovascular disease. At this point, conclusions cannot be drawn. Mercury exposure may also be associated with perturbations in markers of autoimmunity. The clinical significance of these findings remains unclear.

 Heavy metals pose risks to health that are especially burdensome in selected parts of the world. For example, *arsenic* exposure from natural contamination of shallow tube wells inserted for drinking water is a major environmental problem for millions of residents in parts of Bangladesh and Western India. Contamination was formerly considered only a problem with deep wells; however, the geology of this region allows most residents only a few alternatives for potable drinking water. The combustion of leaded gasoline with resulting contamination of air and soil with *lead oxide* remains a problem in some countries of Central Asia, Southeast Asia, Africa, and the Middle East. Populations living in the Arctic have been shown to have particularly high exposures to *mercury* due to long-range transport patterns that concentrate mercury in the polar regions, as well as the traditional dependence of Arctic peoples on the consumption of fish and other wildlife that bioconcentrate methylmercury.

A few additional metals deserve brief mention but are not covered in [Table 449-1](#) because of the relative rarity of their being clinically encountered or the uncertainty regarding their potential toxicities. *Aluminum* contributes to the encephalopathy in patients with severe renal disease, who are undergoing dialysis ([Chap. 403](#)). High levels of aluminum are found in the neurofibrillary tangles in the cerebral cortex and hippocampus of patients with Alzheimer's disease, as well as in the drinking water and soil of areas with an unusually high incidence of Alzheimer's. The experimental and epidemiologic evidence for the aluminum–Alzheimer's disease link remains relatively weak, however, and it cannot be concluded that aluminum is a causal agent or a contributing factor in neurodegenerative disease. Hexavalent chromium is corrosive and sensitizing. Workers in the chromate and chrome pigment production industries have consistently had a greater risk of lung cancer. The introduction of *cobalt* chloride as a fortifier in beer led to outbreaks of fatal cardiomyopathy among heavy consumers. Occupational exposure (e.g., of miners, dry-battery manufacturers, and arc welders) to *manganese* (Mn) can cause a parkinsonian syndrome within 1–2 years, including gait disorders; postural instability; a masked, expressionless face; tremor; and psychiatric symptoms. With the introduction of methylcyclopentadienyl manganese tricarbonyl (MMT) as a gasoline additive, there is concern for the toxic potential of environmental manganese exposure. For example, a recent study found a high prevalence of parkinsonian disorders in a community with risks proportionate to estimated manganese exposures emitted by local ferroalloy industries. Epidemiologic studies have also suggested that manganese may interfere with early childhood neurodevelopment in ways similar to that of lead. Manganese toxicity is clearly associated with dopaminergic dysfunction and its toxicity is likely influenced by age, gender, ethnicity, genetics, and preexisting medical conditions. *Nickel* exposure induces an allergic response, and inhalation of nickel compounds with low aqueous solubility (e.g., nickel subsulfide and nickel oxide) in occupational settings is associated with an increased risk of lung cancer. Overexposure to selenium may cause local irritation of the

respiratory system and eyes, gastrointestinal irritation, liver inflammation, loss of hair, depigmentation, and peripheral nerve damage. Workers exposed to certain organic forms of *tin* (particularly trimethyl and triethyl derivatives) have developed psychomotor disturbances, including tremor, convulsions, hallucinations, and psychotic behavior.

Thallium, which is a component of some insecticides, metal alloys, and fireworks, is absorbed through the skin as well as by ingestion and inhalation. Severe poisoning follows a single ingested dose of >1 g or >8 mg/kg. Nausea and vomiting, abdominal pain, and hematemesis precede confusion, psychosis, organic brain syndrome, and coma. Thallium is radiopaque. Induced emesis or gastric lavage is indicated within 4–6 h of acute ingestion; Prussian blue prevents absorption and is given orally at 250 mg/kg in divided doses. Unlike other types of metal poisoning, thallium poisoning may be less severe when activated charcoal is used to interrupt its enterohepatic circulation. Other measures include forced diuresis, treatment with potassium chloride (which promotes renal excretion of thallium), and peritoneal dialysis.

Chelation Therapy The Trial to Assess Chelation Therapy (TACT), a multi-center double-blind, placebo-controlled, prospective randomized trial funded by NIH of 1708 patients aged ≥50 years who had experienced a myocardial infarction (MI), recently found that a protocol of repeated intravenous chelation with disodium EDTA, compared with placebo, modestly but significantly reduced the risk of adverse cardiovascular outcomes, many of which were revascularization procedures. The effect was particularly pronounced among those with concurrent diabetes. However, the trial did not include rigorous measures of lead exposure or any selection criteria based on lead exposure; thus, even though chelation reduces metal burdens, which have been associated with adverse cardiovascular effects (especially lead), it remains unclear whether the beneficial effects result from a reduction in metal burden. In view of the risks of side effects associated with chelation, by themselves, the results are not sufficient to support the routine use of chelation therapy for treatment of patients either who have had an MI or who have had low-level lead exposure. A follow-up trial with rigorous measures of metals exposure is planned.

FURTHER READING

Alamolhodaie NS et al: Arsenic cardiotoxicity: An overview. *Environ Toxicol Pharmacol* 40:1005, 2015. [\[PubMed: 26606645\]](#)

Aneni EC et al: Chronic toxic metal exposure and cardiovascular disease: Mechanisms of risk and emerging role of chelation therapy. *Curr Atheroscler Rep* 18:81, 2016. [\[PubMed: 27822681\]](#)

Cheng X et al: Cadmium exposure and risk of any fracture: A PRISMA-compliant systematic review and meta-analysis. *Medicine (Baltimore)* 95:e2932, 2016. [\[PubMed: 26962791\]](#)

Gidlow DA: Lead toxicity. *Occup Med (Lond)* 65:348, 2015. [\[PubMed: 26187801\]](#)

Kim KH et al: A review on the distribution of Hg in the environment and its human health impacts. J Hazard Mater 306:376, 2016. [[PubMed: 26826963](#)]

Lucchini RG, Hashim D: Tremor secondary to neurotoxic exposure: Mercury, lead, solvents, pesticides. Handb Clin Neurol 131:241, 2015. [[PubMed: 26563793](#)]

O'Neal SL, Zheng W: Manganese toxicity upon overexposure: A decade in review. Curr Environ Health Rep 2:315, 2015. [[PubMed: 26231508](#)]

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter 450: Poisoning and Drug Overdose

Mark B. Mycyk

INTRODUCTION

Poisoning refers to the development of dose-related adverse effects following exposure to chemicals, drugs, or other xenobiotics. To paraphrase Paracelsus, the dose makes the poison. Although most poisons have predictable dose-related effects, individual responses to a given dose may vary because of genetic polymorphism, enzymatic induction or inhibition in the presence of other xenobiotics, or acquired tolerance. Poisoning may be local (e.g., skin, eyes, or lungs) or systemic depending on the route of exposure, the chemical and physical properties of the poison, and its mechanism of action. The severity and reversibility of poisoning also depend on the functional reserve of the individual or target organ, which is influenced by age and preexisting disease.

EPIDEMIOLOGY

More than 5 million poison exposures occur in the United States each year. Most are acute, are accidental (unintentional), involve a single agent, occur in the home, result in minor or no toxicity, and involve children <6 years of age. Pharmaceuticals are involved in 47% of exposures and in 84% of serious or fatal poisonings. In the last decade, the rate of injury-related deaths from poisoning has overtaken the rate of deaths related to motor-vehicle crashes in the United States. According to the Centers for Disease Control (CDC), twice as many Americans died from drug overdoses in 2014 compared to 2000. Although prescription opioids have appropriately received attention as a major reason for the increased number of poisoning deaths, the availability of other pharmaceuticals and rapid proliferation of novel drugs of abuse also contribute to the increasing death rate. In many parts of the United States, where these issues are particularly prevalent, there are efforts to develop better prescription drug databases and enhanced training for health care professionals in pain management and the use of opiates. Unintentional exposures can result from the improper use of chemicals at work or play; label misreading; product mislabeling; mistaken identification of unlabeled chemicals; uninformed self-medication; and dosing errors by nurses, pharmacists, physicians, parents, and the elderly. Excluding the recreational use of ethanol, attempted suicide (deliberate self-harm) is the most common reported reason for intentional poisoning. Recreational use of prescribed and over-the-counter drugs for psychotropic or euphoric effects (*abuse*) or excessive self-dosing (*misuse*) is increasingly common and may also result in unintentional self-poisoning.

About 20–25% of exposures require bedside health-professional evaluation, and 5% of all exposures require hospitalization. Poisonings account for 5–10% of all ambulance transports, emergency department visits, and

intensive care unit admissions. Hospital admissions related to poisoning are also associated with longer lengths of stay and increase the utilization of resources such as radiography and other laboratory services. Up to 30% of psychiatric admissions are prompted by attempted suicide via overdosage. Overall, the mortality rate is low: <1% of all poisoning exposures. It is significantly higher (1–2%) among hospitalized patients with intentional (suicidal) overdose or complications from drugs of abuse, who account for the majority of serious poisonings. Acetaminophen is the pharmaceutical agent most often implicated in fatal poisoning. Overall, carbon monoxide is the leading cause of death from poisoning, but this prominence is not reflected in hospital or poison center statistics because patients with such poisoning are typically dead when discovered and are referred directly to medical examiners.

DIAGNOSIS

Although poisoning can mimic other illnesses, the correct diagnosis can usually be established by the history, physical examination, routine and toxicologic laboratory evaluations, and characteristic clinical course.

HISTORY

The *history* should include the time, route, duration, and circumstances (location, surrounding events, and intent) of exposure; the name and amount of each drug, chemical, or ingredient involved; the time of onset, nature, and severity of symptoms; the time and type of first-aid measures provided; and the medical and psychiatric history.

In many cases the patient is confused, comatose, unaware of an exposure, or unable or unwilling to admit to one. Suspicious circumstances include unexplained sudden illness in a previously healthy person or a group of healthy people; a history of psychiatric problems (particularly depression); recent changes in health, economic status, or social relationships; and onset of illness during work with chemicals or after ingestion of food, drink (especially ethanol), or medications. When patients become ill soon after arriving from a foreign country or being arrested for criminal activity, “body packing” or “body stuffing” (ingesting or concealing illicit drugs in a body cavity) should be suspected. Relevant information may be available from family, friends, paramedics, police, pharmacists, physicians, and employers, who should be questioned regarding the patient’s habits, hobbies, behavioral changes, available medications, and antecedent events. Patients need to be asked explicitly about their prescribed medications and recreational drug use. Drugs previously considered “illicit” such as cannabinoids are now legal in many places and prescribed for therapeutic purposes. A search of clothes, belongings, and place of discovery may reveal a suicide note or a container of drugs or chemicals. Without a clear history in a patient clinically suspected to be poisoned, all medications available anywhere in the patient’s home or belongings should be considered as possible agents, including medications for pets. The imprint code on pills and the label on chemical products may be used to identify the ingredients and potential toxicity of a suspected poison by consulting a reference text, a computerized database, the manufacturer, or a regional poison information center (800-222-1222). Occupational exposures require review of any available safety data sheet (SDS) from the worksite. Because of increasing globalization from travel and internet consumerism, unfamiliar poisonings may result in local emergency department evaluation. Pharmaceuticals, industrial

chemicals, or drugs of abuse from foreign countries may be identified with the assistance of a regional poison center or via the World Wide Web.

PHYSICAL EXAMINATION AND CLINICAL COURSE

The *physical examination* should focus initially on vital signs, the cardiopulmonary system, and neurologic status. The neurologic examination should include documentation of neuromuscular abnormalities such as dyskinesia, dystonia, fasciculations, myoclonus, rigidity, and tremors. The patient should also be examined for evidence of trauma and underlying illnesses. Focal neurologic findings are uncommon in poisoning, and their presence should prompt evaluation for a structural central nervous system (CNS) lesion. Examination of the eyes (for nystagmus and pupil size and reactivity), abdomen (for bowel activity and bladder size), and skin (for burns, bullae, color, warmth, moisture, pressure sores, and puncture marks) may reveal findings of diagnostic value. When the history is unclear, all orifices should be examined for the presence of chemical burns and drug packets. The odor of breath or vomitus and the color of nails, skin, or urine may provide important diagnostic clues.

The diagnosis of poisoning in cases of unknown etiology primarily relies on pattern recognition. The first step is to assess the pulse, blood pressure, respiratory rate, temperature, and neurologic status and to characterize the overall physiologic state as stimulated, depressed, discordant, or normal (**Table 450-1**). Obtaining a complete set of vital signs and reassessing them frequently are critical. Measuring core temperature is especially important, even in difficult or combative patients, since temperature elevation is the most reliable prognosticator of poor outcome in poisoning from stimulants (e.g., cocaine) or drug withdrawal (e.g., alcohol or GHB). The next step is to consider the underlying causes of the physiologic state and to attempt to identify a pathophysiologic pattern or toxic syndrome (*toxidrome*) based on the observed findings. Assessing the severity of physiologic derangements (**Table 450-2**) is useful in this regard and also for monitoring the clinical course and response to treatment. The final step is to attempt to identify the particular agent involved by looking for unique or relatively poison-specific physical or ancillary test abnormalities. Distinguishing among toxidromes on the basis of the physiologic state is summarized next.

TABLE 450-1

Differential Diagnosis of Poisoning Based on Physiologic State

STIMULATED	DEPRESSED	DISCORDANT	NORMAL
Sympathetics Sympathomimetics Ergot alkaloids Methylxanthines Monoamine oxidase inhibitors Thyroid hormones Anticholinergics Antihistamines Antiparkinsonian agents Antipsychotics Antispasmodics Belladonna alkaloids Cyclic antidepressants Mushrooms and plants Hallucinogens Cannabinoids (marijuana) LSD and analogues Mescaline and analogues Mushrooms Phencyclidine and analogues Withdrawal syndromes Barbiturates Benzodiazepines	Sympatholytics α_1 -Adrenergic antagonists α_2 -Adrenergic agonists ACE inhibitors Angiotensin receptor blockers Antipsychotics β -Adrenergic blockers Calcium channel blockers Cardiac glycosides Cyclic antidepressants Cholinergics Acetylcholinesterase inhibitors Muscarinic agonists Nicotinic agonists Opioids Analgesics GI antispasmodics Heroin Sedative-hypnotics Alcohols Anticonvulsants Barbiturates Benzodiazepines GABA precursors Muscle relaxants	Asphyxiants Cytochrome oxidase inhibitors Inert gases Irritant gases Methemoglobin inducers Oxidative phosphorylation inhibitors AGMA inducers Alcohol (ketoacidosis) Ethylene glycol Iron Methanol Other alcohols Salicylate Toluene CNS syndromes Extrapyramidal reactions Hydrocarbon inhalation Isoniazid Lithium Neuroleptic malignant syndrome Serotonin syndrome Strychnine Membrane-active agents Amantadine Antiarrhythmics Antihistamines	Nontoxic exposure Psychogenic illness “Toxic time-bombs” Slow absorption Anticholinergics Carbamazepine Concretion formers Extended-release phenytoin sodium capsules (Dilantin Kapseals) Drug packets Enteric-coated pills Diphenoxylate-atropine (Lomotil) Opioids Salicylates Sustained-release pills Valproate Slow distribution Cardiac glycosides Lithium Metals Salicylate Valproate Toxic metabolite Acetaminophen Carbon tetrachloride Cyanogenic glycosides Ethylene glycol Methanol Methemoglobin inducers Mushroom toxins Organophosphate insecticides Paraquat Metabolism disruptors Antineoplastic agents Antiviral agents Colchicine

Ethanol STIMULATED GHB products	Other agents DEPRESSED GHB products	DISCORDANT	Hypoglycemic agents NORMAL Immunosuppressive agents
Opioids Sedative- hypnotics Sympatholytics		Antipsychotics Carbamazepine Cyclic antidepressants Local anesthetics Opioids (some) Quinoline antimalarials	MAO inhibitors Metals Other oral anticoagulants Salicylate Warfarin

Abbreviations: ACE, angiotensin-converting enzyme; AGMA, anion-gap metabolic acidosis; CNS, central nervous system; GABA, γ -aminobutyric acid; GHB, γ -hydroxybutyrate; GI, gastrointestinal; LSD, lysergic acid diethylamide; MAO, monoamine oxidase.

TABLE 450-2

Severity of Physiologic Stimulation and Depression in Poisoning and Drug Withdrawal

Physiologic Stimulation	
Grade 1	Anxious, irritable, tremulous; vital signs normal; diaphoresis, flushing or pallor, mydriasis, and hyperreflexia sometimes present
Grade 2	Agitated; may have confusion or hallucinations but can converse and follow commands; vital signs mildly to moderately increased
Grade 3	Delirious; unintelligible speech, uncontrollable motor hyperactivity; moderately to markedly increased vital signs; tachyarrhythmias possible
Grade 4	Coma, seizures, cardiovascular collapse
Physiologic Depression	
Grade 1	Awake, lethargic, or sleeping but arousable by voice or tactile stimulation; able to converse and follow commands; may be confused
Grade 2	Responds to pain but not voice; can vocalize but not converse; spontaneous motor activity present; brainstem reflexes intact
Grade 3	Unresponsive to pain; spontaneous motor activity absent; brainstem reflexes depressed; motor tone, respirations, and temperature decreased
Grade 4	Unresponsive to pain; flaccid paralysis; brainstem reflexes and respirations absent; cardiovascular vital signs decreased

The Stimulated Physiologic State

Increased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity characterize the *stimulated* physiologic state, which can reflect sympathetic, anticholinergic, or hallucinogen poisoning or drug withdrawal (Table 450-1). Other features are noted in (Table 450-2). Mydriasis, a characteristic feature of all stimulants, is most marked in anticholinergic poisoning since pupillary reactivity relies on muscarinic control. In sympathetic poisoning (e.g., due to cocaine), pupils are also enlarged, but some reactivity to light remains. The anticholinergic toxidrome is also distinguished by hot, dry, flushed skin; decreased bowel sounds; and urinary retention. Other stimulant syndromes increase sympathetic activity and cause diaphoresis, pallor, and increased bowel activity with varying degrees of nausea, vomiting, abnormal distress, and occasionally diarrhea. The absolute and relative degree of vital-sign changes and neuromuscular hyperactivity can help distinguish among stimulant toxidromes. Since sympathetics stimulate the peripheral nervous system more directly than do

hallucinogens or drug withdrawal, markedly increased vital signs and organ ischemia suggest sympathetic poisoning. Findings helpful in suggesting the particular drug or class causing physiologic stimulation include reflex bradycardia from selective α -adrenergic stimulants (e.g., decongestants), hypotension from selective β -adrenergic stimulants (e.g., asthma therapeutics), limb ischemia from ergot alkaloids, rotatory nystagmus from phencyclidine and ketamine (the only physiologic stimulants that cause this finding), and delayed cardiac conduction from high doses of cocaine and some anticholinergic agents (e.g., antihistamines, cyclic antidepressants, and antipsychotics). Seizures suggest a sympathetic etiology, an anticholinergic agent with membrane-active properties (e.g., cyclic antidepressants, phenothiazines), or a withdrawal syndrome. Close attention to core temperature is critical in patients with grade 4 physiologic stimulation ([Table 450-2](#)).

The Depressed Physiologic State

Decreased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity are indicative of the *depressed* physiologic state caused by “functional” sympatholytics (agents that decrease cardiac function and vascular tone as well as sympathetic activity), cholinergic (muscarinic and nicotinic) agents, opioids, and sedative-hypnotic γ -aminobutyric acid (GABA)-ergic agents ([Table 450-1](#) and [450-2](#)). Miosis is also common and is most pronounced in opioid and cholinergic poisoning. Miosis is distinguished from other depressant syndromes by muscarinic and nicotinic signs and symptoms ([Table 450-1](#)). Pronounced cardiovascular depression in the absence of significant CNS depression suggests a direct or peripherally acting sympatholytic. In contrast, in opioid and sedative-hypnotic poisoning, vital-sign changes are secondary to depression of CNS cardiovascular and respiratory centers (or consequent hypoxemia), and significant abnormalities in these parameters do not occur until there is a marked decrease in the level of consciousness (grade 3 or 4 physiologic depression; [[Table 450-2](#)]). Other clues that suggest the cause of physiologic depression include cardiac arrhythmias and conduction disturbances (due to antiarrhythmics, β -adrenergic antagonists, calcium channel blockers, digitalis glycosides, propoxyphene, and cyclic antidepressants), mydriasis (due to tricyclic antidepressants, some antiarrhythmics, meperidine, and diphenoxylate-atropine [Lomotil]), nystagmus (due to sedative-hypnotics), and seizures (due to cholinergic agents, propoxyphene, and cyclic antidepressants).

The Discordant Physiologic State

The *discordant* physiologic state is characterized by mixed vital-sign and neuromuscular abnormalities, as observed in poisoning by asphyxiants, CNS syndromes, membrane-active agents, and anion-gap metabolic acidosis (AGMA) inducers ([Table 450-1](#)). In these conditions, manifestations of physiologic stimulation and physiologic depression occur together or at different times during the clinical course. For example, membrane-active agents can cause simultaneous coma, seizures, hypotension, and tachyarrhythmias. Alternatively, vital signs may be normal while the patient has an altered mental status or is obviously sick or clearly symptomatic. Early, pronounced vital-sign and mental-status changes suggest asphyxiant or membrane-active agent poisoning; the lack of such abnormalities suggests an AGMA inducer; and marked neuromuscular dysfunction without significant vital-sign abnormalities suggests a CNS syndrome. The *discordant* physiologic state may also be evident in patients poisoned with multiple agents.

The Normal Physiologic State

A *normal* physiologic status and physical examination may be due to a nontoxic exposure, psychogenic illness, or poisoning by “toxic time-bombs”: agents that are slowly absorbed, are slowly distributed to their sites of action, require metabolic activation, or disrupt metabolic processes ([Table 450-1](#)). Because so many medications have now been reformulated into a once-a-day preparations for the patient’s convenience and adherence, toxic

time-bombs are increasingly common. Diagnosing a nontoxic exposure requires that the identity of the exposure agent be known or that a toxic time-bomb exposure be excluded and the time since exposure exceed the longest known or predicted interval between exposure and peak toxicity. Psychogenic illness (fear of being poisoned, mass hysteria) may also follow a nontoxic exposure and should be considered when symptoms are inconsistent with exposure history. Anxiety reactions resulting from a nontoxic exposure can cause mild physiologic stimulation ([Table 450-2](#)) and be indistinguishable from toxicologic causes without ancillary testing or a suitable period of observation.

LABORATORY ASSESSMENT

Laboratory assessment may be helpful in the differential diagnosis. Increased anion gap metabolic acidosis (AGMA) is most common in advanced methanol, ethylene glycol, and salicylate intoxication but can occur with any poisoning that results in hepatic, renal, or respiratory failure; seizures; or shock. The serum lactate concentration is more commonly low (less than the anion gap) in the former and high (nearly equal to the anion gap) in the latter. An abnormally low anion gap can be due to elevated blood levels of bromide, calcium, iodine, lithium, or magnesium. An increased osmolal gap—a difference of >10 mmol/L between serum osmolality (measured by freezing-point depression) and osmolality calculated from serum sodium, glucose, and blood [urea](#) nitrogen levels—suggests the presence of a low-molecular-weight solute such as acetone; an alcohol (benzyl, ethanol, isopropanol, methanol); a glycol (diethylene, ethylene, propylene); ether (ethyl, glycol); or an “unmeasured” cation (calcium, magnesium) or sugar (glycerol, [mannitol](#), sorbitol). Ketosis suggests acetone, isopropyl alcohol, salicylate poisoning, or alcoholic ketoacidosis. Hypoglycemia may be due to poisoning with β -adrenergic blockers, ethanol, insulin, oral hypoglycemic agents, quinine, and salicylates, whereas hyperglycemia can occur in poisoning with acetone, β -adrenergic agonists, caffeine, calcium channel blockers, iron, [theophylline](#), or *N*-3-pyridylmethyl-*N'*-*p*-nitrophenylurea (PNU [Vacor]). Hypokalemia can be caused by barium, β -adrenergic agonists, caffeine, diuretics, [theophylline](#), or toluene; hyperkalemia suggests poisoning with an α -adrenergic agonist, a β -adrenergic blocker, cardiac glycosides, or fluoride. Hypocalcemia may be seen in ethylene glycol, fluoride, and oxalate poisoning. PT and INR are useful for risk stratification in cases of warfarin or rodenticide poisoning, but are not to be relied on when evaluating overdose or complications from new anticoagulant pharmaceuticals (e.g., dabigatran).

The *electrocardiogram* (ECG) can be useful for rapid diagnostic purposes. Bradycardia and atrioventricular block may occur in patients poisoned by α -adrenergic agonists, antiarrhythmic agents, beta blockers, calcium channel blockers, cholinergic agents (carbamate and organophosphate insecticides), cardiac glycosides, lithium, or tricyclic antidepressants. QRS- and QT-interval prolongation may be caused by hyperkalemia, various antidepressants, and other membrane-active drugs ([Table 450-1](#)). Ventricular tachyarrhythmias may be seen in poisoning with cardiac glycosides, fluorides, membrane-active drugs, methylxanthines, sympathomimetics, antidepressants, and agents that cause hyperkalemia or potentiate the effects of endogenous catecholamines (e.g., chloral hydrate, aliphatic and halogenated hydrocarbons).

Radiologic studies may occasionally be useful. Pulmonary edema (adult respiratory distress syndrome [ARDS]) can be caused by poisoning with carbon monoxide, cyanide, an opioid, paraquat, phencyclidine, a sedative-hypnotic, or salicylate; by inhalation of irritant gases, fumes, or vapors (acids and alkali, ammonia, aldehydes, chlorine, hydrogen sulfide, isocyanates, metal oxides, mercury, phosgene, polymers); or by prolonged anoxia,

hyperthermia, or shock. Aspiration pneumonia is common in patients with coma, seizures, and petroleum distillate aspiration. Chest x-ray is useful for identifying complications from metal fume fever or elemental mercury. The presence of radiopaque densities on abdominal x-rays or abdominal CT scan suggests the ingestion of calcium salts, chloral hydrate, chlorinated hydrocarbons, heavy metals, illicit drug packets, iodinated compounds, potassium salts, enteric-coated tablets, or salicylates.

Toxicologic analysis of urine and blood (and occasionally of gastric contents and chemical samples) can sometimes confirm or rule out suspected poisoning. Interpretation of laboratory data requires knowledge of the qualitative and quantitative tests used for screening and confirmation (enzyme-multiplied, fluorescence polarization, and radio-immunoassays; colorimetric and fluorometric assays; thin-layer, gas-liquid, or high-performance liquid chromatography; gas chromatography; mass spectrometry), their sensitivity (limit of detection) and specificity, the preferred biologic specimen for analysis, and the optimal time of specimen sampling. Personal communication with the hospital laboratory is essential to an understanding of institutional testing capabilities and limitations.

Rapid *qualitative* hospital-based urine tests for drugs of abuse are only screening tests that cannot confirm the exact identity of the detected substance and should not be considered diagnostic or used for forensic purposes: False-positive and false-negative results are common. A positive screen may result from other pharmaceuticals that interfere with laboratory analysis (e.g., fluoroquinolones commonly cause “false-positive” opiate screens). Confirmatory testing with gas chromatography/mass spectrometry can be requested, but it often takes weeks to obtain a reported result. A negative screening result may mean that the responsible substance is not detectable by the test used or that its concentration is too low for detection at the time of sampling. For instance, recent new drugs of abuse that often result in emergency department evaluation for unexpected complications, such as synthetic cannabinoids (spice), cathinones (bath salts), and opiate substitutes (kratom), are not detectable by hospital-based tests. In cases where a drug concentration is too low to be detected early during clinical evaluation, repeating the test at a later time may yield a positive result. Patients symptomatic from drugs of abuse often require immediate management based on the history, physical examination, and observed toxidrome without laboratory confirmation (e.g., apnea from opioid intoxication). When the patient is asymptomatic or when the clinical picture is consistent with the reported history, qualitative screening is neither clinically useful nor cost-effective. Thus, qualitative drug screens are of greatest value for the evaluation of patients with severe or unexplained toxicities, such as coma, seizures, cardiovascular instability, metabolic or respiratory acidosis, and nonsinus cardiac rhythms. In contrast to qualitative drug screens, *quantitative* serum tests are useful for evaluation of patients poisoned with acetaminophen (**Chap. 333**), alcohols (including ethylene glycol and methanol), anticonvulsants, barbiturates, digoxin, heavy metals, iron, lithium, salicylate, and **theophylline** as well as for the presence of carboxyhemoglobin and methemoglobin. The serum concentration in these cases guides clinical management, and results are often available within an hour.

The *response to antidotes* is sometimes useful for diagnostic purposes. Resolution of altered mental status and abnormal vital signs within minutes of IV administration of **dextrose**, naloxone, or flumazenil is virtually diagnostic of hypoglycemia, opioid poisoning, and benzodiazepine intoxication, respectively. The prompt reversal of dystonic (extrapyramidal) signs and symptoms following an IV dose of benztropine or **diphenhydramine** confirms a drug etiology. Although complete reversal of both central and peripheral

manifestations of anticholinergic poisoning by physostigmine is diagnostic of this condition, physostigmine may cause some arousal in patients with CNS depression of any etiology.

TREATMENT

TREATMENT

Poisoning and Drug Overdose

GENERAL PRINCIPLES

Treatment goals include support of vital signs, prevention of further poison absorption (decontamination), enhancement of poison elimination, administration of specific antidotes, and prevention of reexposure ([Table 450-3](#)). Specific treatment depends on the identity of the poison, the route and amount of exposure, the time of presentation relative to the time of exposure, and the severity of poisoning. Knowledge of the offending agents' pharmacokinetics and pharmacodynamics is essential.

During the *pretoxic phase*, prior to the onset of poisoning, decontamination is the highest priority, and treatment is based solely on the history. The maximal potential toxicity based on the greatest possible exposure should be assumed. Since decontamination is more effective when accomplished soon after exposure and when the patient is asymptomatic, the initial history and physical examination should be focused and brief. It is also advisable to establish IV access and initiate cardiac monitoring, particularly in patients with potentially serious ingestions or unclear histories.

When an accurate history is not obtainable and a poison causing delayed toxicity (i.e., a toxic time-bomb) or irreversible damage is suspected, blood and urine should be sent for appropriate toxicologic screening and quantitative analysis. During poison absorption and distribution, blood levels may be greater than those in tissue and may not correlate with toxicity. However, high blood levels of agents whose metabolites are more toxic than the parent compound (acetaminophen, ethylene glycol, or methanol) may indicate the need for additional interventions (antidotes, dialysis). Most patients who remain asymptomatic or who become asymptomatic 6 h after ingestion are unlikely to develop subsequent toxicity and can be discharged safely. Longer observation will be necessary for patients who have ingested toxic time-bombs.

During the *toxic phase*—the interval between the onset of poisoning and its peak effects—management is based primarily on clinical and laboratory findings. *Effects after an overdose usually begin sooner, peak later, and last longer than they do after a therapeutic dose.* A drug's published pharmacokinetic profile in standard references such as the *Physician's Desk Reference* (PDR) is usually different from its toxicokinetic profile in overdose. Resuscitation and stabilization are the first priority. Symptomatic patients should have an IV line placed and should undergo oxygen saturation determination, cardiac monitoring, and continuous observation. Baseline laboratory, ECG, and x-ray evaluation may also be appropriate. Intravenous glucose (unless the serum level is documented to be normal), naloxone, and [thiamine](#) should be considered in patients with altered mental status, particularly those with coma or seizures. Decontamination should also be considered, but it is less likely to be effective during this phase than during the *pretoxic phase*.

Measures that enhance poison elimination may shorten the duration and severity of the toxic phase. However, they are not without risk, which must be weighed against the potential benefit. Diagnostic certainty (usually via laboratory confirmation) is generally a prerequisite. Intestinal (gut) dialysis with repetitive doses of activated charcoal (see “[Multiple-Dose Activated Charcoal](#),” later) can enhance the elimination of selected poisons such as [theophylline](#) or carbamazepine. Urinary alkalization may enhance the elimination of salicylates and a few other poisons. Chelation therapy can enhance the elimination of selected metals. Extracorporeal elimination methods are effective for many poisons, but their expense and risk make their use reasonable only in patients who would otherwise have an unfavorable outcome.

During the *resolution phase* of poisoning, supportive care and monitoring should continue until clinical, laboratory, and ECG abnormalities have resolved. Since chemicals are eliminated sooner from the blood than from tissues, blood levels are usually lower than tissue levels during this phase and again may not correlate with toxicity. This discrepancy applies particularly when extracorporeal elimination procedures are used. Redistribution from tissues may cause a rebound increase in the blood level after termination of these procedures (e.g., lithium). When a metabolite is responsible for toxic effects, continued treatment may be necessary in the absence of clinical toxicity or abnormal laboratory studies.

SUPPORTIVE CARE

The goal of supportive therapy is to maintain physiologic homeostasis until detoxification is accomplished and to prevent and treat secondary complications such as aspiration, bedsores, cerebral and pulmonary edema, pneumonia, rhabdomyolysis, renal failure, sepsis, thromboembolic disease, coagulopathy, and generalized organ dysfunction due to hypoxemia or shock.

Admission to an intensive care unit is indicated for the following: patients with severe poisoning (coma, respiratory depression, hypotension, cardiac conduction abnormalities, cardiac arrhythmias, hypothermia or hyperthermia, seizures); those needing close monitoring, antidotes, or enhanced elimination therapy; those showing progressive clinical deterioration; and those with significant underlying medical problems. Patients with mild to moderate toxicity can be managed on a general medical service, on an intermediate care unit, or in an emergency department observation area, depending on the anticipated duration and level of monitoring needed (intermittent clinical observation versus continuous clinical, cardiac, and respiratory monitoring). Patients who have attempted suicide require continuous observation and measures to prevent self-injury until they are no longer suicidal.

Respiratory Care

Endotracheal intubation for protection against the aspiration of gastrointestinal contents is of paramount importance in patients with CNS depression or seizures as this complication can increase morbidity and mortality rates. Mechanical ventilation may be necessary for patients with respiratory depression or hypoxemia and for facilitation of therapeutic sedation or paralysis of patients in order to prevent or treat hyperthermia, acidosis, and rhabdomyolysis associated with neuromuscular hyperactivity. Since clinical assessment of respiratory function can be inaccurate, the need for oxygenation and ventilation is best determined by continuous pulse oximetry or arterial blood-gas analysis. The gag reflex is not a reliable indicator of the need for intubation. A patient with CNS depression may maintain airway patency while being stimulated but not if left alone. Drug-induced pulmonary edema is usually noncardiac rather than cardiac in origin, although profound

CNS depression and cardiac conduction abnormalities suggest the latter. Measurement of pulmonary artery pressure may be necessary to establish the cause and direct appropriate therapy. Extracorporeal measures (membrane oxygenation, ECMO, venoarterial perfusion, cardiopulmonary bypass) and partial liquid (perfluorocarbon) ventilation may be appropriate for severe but reversible respiratory failure.

Cardiovascular Therapy

Maintenance of normal tissue perfusion is critical for complete recovery to occur once the offending agent has been eliminated. Focused bedside echocardiography or measurement of CVP may help prioritize therapeutic strategies. If hypotension is unresponsive to volume expansion and appropriate goal-directed antidotal therapy, treatment with norepinephrine, [epinephrine](#), or high-dose dopamine may be necessary. Intraaortic balloon pump counterpulsation and venoarterial or cardiopulmonary perfusion techniques should be considered for severe but reversible cardiac failure. For patients with a return of spontaneous circulation after resuscitative treatment for cardiopulmonary arrest secondary to poisoning, therapeutic hypothermia should be used according to protocol. Bradyarrhythmias associated with hypotension generally should be treated as described in [Chaps. 239 and 240](#). Glucagon, calcium, and high-dose insulin with [dextrose](#) may be effective in beta blocker and calcium channel blocker poisoning. Antibody therapy may be indicated for cardiac glycoside poisoning.

Supraventricular tachycardia associated with hypertension and CNS excitation is almost always due to agents that cause generalized physiologic excitation ([Table 450–1](#)). Most cases are mild or moderate in severity and require only observation or nonspecific sedation with a benzodiazepine. In severe cases or those associated with hemodynamic instability, chest pain, or ECG evidence of ischemia, specific therapy is indicated. When the etiology is sympathetic hyperactivity, treatment with a benzodiazepine should be prioritized. Further treatment with a combined alpha and beta blocker ([labetalol](#)), a calcium channel blocker ([verapamil](#) or diltiazem), or a combination of a beta blocker and a vasodilator (esmolol and nitroprusside) may be considered for cases refractory to high doses of benzodiazepines only when adequate sedation has been achieved but cardiac conduction or blood pressure abnormalities persist. Treatment with an α -adrenergic antagonist (phentolamine) alone may sometimes be appropriate. If the cause is anticholinergic poisoning, physostigmine alone can be effective. Supraventricular tachycardia without hypertension is generally secondary to vasodilation or hypovolemia and responds to fluid administration.

For ventricular tachyarrhythmias due to tricyclic antidepressants and other membrane-active agents ([Table 450-1](#)), [sodium bicarbonate](#) is indicated, whereas class IA, IC, and III antiarrhythmic agents are contraindicated because of similar electrophysiologic effects. Although [lidocaine](#) and phenytoin are historically safe for ventricular tachyarrhythmias of any etiology, [sodium bicarbonate](#) should be considered first for any ventricular arrhythmia suspected to have a toxicologic etiology. Intravenous lipid emulsion therapy has shown benefit for treatment of arrhythmias and hemodynamic instability from various membrane-active agents. Beta blockers can be hazardous if the arrhythmia is due to sympathetic hyperactivity. Magnesium sulfate and overdrive pacing (by isoproterenol or a pacemaker) may be useful in patients with torsades des pointes and prolonged QT intervals. Magnesium and anti-digoxin antibodies should be considered in patients with severe cardiac glycoside poisoning. Invasive (esophageal or intracardiac) ECG recording may be necessary to determine the origin (ventricular or supraventricular) of wide-complex tachycardias ([Chap. 241](#)). If the patient is hemodynamically stable, however, it is reasonable to simply observe him or her rather than to administer another potentially

proarrhythmic agent. Arrhythmias may be resistant to drug therapy until underlying acid-base, electrolyte, oxygenation, and temperature derangements are corrected.

Central Nervous System Therapies

Neuromuscular hyperactivity and seizures can lead to hyperthermia, lactic acidosis, and rhabdomyolysis and should be treated aggressively. Seizures caused by excessive stimulation of catecholamine receptors (sympathomimetic or hallucinogen poisoning and drug withdrawal) or decreased activity of GABA ([isoniazid](#) poisoning) or glycine (strychnine poisoning) receptors are best treated with agents that enhance GABA activity, such as benzodiazepine or barbiturates. Since benzodiazepines and barbiturates act by slightly different mechanisms (the former increases the frequency via allosteric modulation at the receptor and the latter directly increases the duration of chloride channel opening in response to GABA), therapy with both may be effective when neither is effective alone. Seizures caused by [isoniazid](#), which inhibits the synthesis of GABA at several steps by interfering with the cofactor [pyridoxine](#) (vitamin B₆), may require high doses of supplemental [pyridoxine](#). Seizures resulting from membrane destabilization (beta blocker or cyclic antidepressant poisoning) require GABA enhancers (benzodiazepines first, barbiturates second). Phenytoin is contraindicated in toxicologic seizures: Animal and human data demonstrate worse outcomes after phenytoin loading, especially in [theophylline](#) overdose. For poisons with central dopaminergic effects (methamphetamine, phencyclidine) manifested by psychotic behavior, a dopamine receptor antagonist, such as haloperidol or ziprasidone, may be useful. In anticholinergic and cyanide poisoning, specific antidotal therapy may be necessary. The treatment of seizures secondary to cerebral ischemia or edema or to metabolic abnormalities should include correction of the underlying cause. Neuromuscular paralysis is indicated in refractory cases. Electroencephalographic monitoring and continuing treatment of seizures are necessary to prevent permanent neurologic damage. Serotonergic receptor overstimulation in serotonin syndrome may be treated with cyproheptadine.

Other Measures

Temperature extremes, metabolic abnormalities, hepatic and renal dysfunction, and secondary complications should be treated by standard therapies.

PREVENTION OF POISON ABSORPTION

Gastrointestinal Decontamination

Whether or not to perform gastrointestinal decontamination and which procedure to use depends on the time since ingestion; the existing and predicted toxicity of the ingestant; the availability, efficacy, and contraindications of the procedure; and the nature, severity, and risk of complications. The efficacy of all decontamination procedures decreases with time, and data are insufficient to support or exclude a beneficial effect when they are used >1 h after ingestion. The average time from ingestion to presentation for treatment is >1 h for children and >3 h for adults. Most patients will recover from poisoning uneventfully with good supportive care alone, but complications of gastrointestinal decontamination, particularly aspiration, can prolong this process. Hence, gastrointestinal decontamination should be performed selectively, not routinely, in the management of overdose patients. It is clearly unnecessary when predicted toxicity is minimal or the time of expected maximal toxicity has passed without significant effect.

Activated charcoal has comparable or greater efficacy; has fewer contraindications and complications; and is less aversive and invasive than ipecac or gastric lavage. Thus it is the preferred method of gastrointestinal decontamination in most situations. Activated charcoal suspension (in water) is given orally via a cup, straw, or small-bore nasogastric tube. The generally recommended dose is 1 g/kg body weight because of its dosing convenience, although in vitro and in vivo studies have demonstrated that charcoal adsorbs $\geq 90\%$ of most substances when given in an amount equal to 10 times the weight of the substance. Palatability may be increased by adding a sweetener (sorbitol) or a flavoring agent (cherry, chocolate, or cola syrup) to the suspension. Charcoal adsorbs ingested poisons within the gut lumen, allowing the charcoal-toxin complex to be evacuated with stool. Charged (ionized) chemicals such as mineral acids, alkalis, and highly dissociated salts of cyanide, fluoride, iron, lithium, and other inorganic compounds are not well adsorbed by charcoal. In studies with animals and human volunteers, charcoal decreases the absorption of ingestants by an average of 73% when given within 5 min of ingestant administration, 51% when given at 30 min, and 36% when given at 60 min. For this reason, charcoal given before hospital arrival increases the potential clinical benefit. Side effects of charcoal include nausea, vomiting, and diarrhea or constipation. Charcoal may also prevent the absorption of orally administered therapeutic agents. Complications include mechanical obstruction of the airway, aspiration, vomiting, and bowel obstruction and infarction caused by inspissated charcoal. Charcoal is not recommended for patients who have ingested corrosives because it obscures endoscopy.

Gastric lavage should be considered for life-threatening poisons that cannot be treated effectively with other decontamination, elimination, or antidotal therapies (e.g., colchicine). Gastric lavage is performed by sequentially administering and aspirating ~5 mL of fluid per kilogram of body weight through a no. 40 French orogastric tube (no. 28 French tube for children). Except in infants, for whom normal saline is recommended, tap water is acceptable. The patient should be placed in Trendelenburg and left lateral decubitus positions to prevent aspiration (even if an endotracheal tube is in place). Lavage decreases ingestant absorption by an average of 52% if performed within 5 min of ingestion administration, 26% if performed at 30 min, and 16% if performed at 60 min. Significant amounts of ingested drug are recovered from <10% of patients. Aspiration is a common complication (occurring in up to 10% of patients), especially when lavage is performed improperly. Serious complications (esophageal and gastric perforation, tube misplacement in the trachea) occur in ~1% of patients. For this reason, the physician should personally insert the lavage tube and confirm its placement, and the patient must be cooperative during the procedure. Gastric lavage is contraindicated in corrosive or petroleum distillate ingestions because of the respective risks of gastroesophageal perforation and aspiration pneumonitis. It is also contraindicated in patients with a compromised unprotected airway and those at risk for hemorrhage or perforation due to esophageal or gastric pathology or recent surgery. Finally, gastric lavage is absolutely contraindicated in combative patients or those who refuse, as most published complications involve patient resistance to the procedure.

Syrup of ipecac, an emetogenic agent that was once the substance most commonly used for decontamination, no longer has a role in poisoning management. Even the American Academy of Pediatrics—traditionally the strongest proponent of ipecac—issued a policy statement in 2003 recommending that ipecac should no longer be used in poisoning treatment. Chronic ipecac use (by patients with anorexia nervosa or bulimia) has been reported to cause electrolyte and fluid abnormalities, cardiac toxicity, and myopathy.

Whole-bowel irrigation is performed by administering a bowel-cleansing solution containing electrolytes and polyethylene glycol (Golytely, Colyte) orally or by gastric tube at a rate of 2 L/h (0.5 L/h in children) until rectal effluent is clear. The patient must be in a sitting position. Although data are limited, whole-bowel irrigation appears to be as effective as other decontamination procedures in volunteer studies. It is most appropriate for those who have ingested foreign bodies, packets of illicit drugs, and agents that are poorly adsorbed by charcoal (e.g., heavy metals). This procedure is contraindicated in patients with bowel obstruction, ileus, hemodynamic instability, and compromised unprotected airways.

Cathartics are salts (disodium phosphate, magnesium citrate and sulfate, sodium sulfate) or saccharides (mannitol, sorbitol) that historically have been given with activated charcoal to promote the rectal evacuation of gastrointestinal contents. However, no animal, volunteer, or clinical data have ever demonstrated any decontamination benefit from cathartics. Abdominal cramps, nausea, and occasional vomiting are side effects. Complications of repeated dosing include severe electrolyte disturbances and excessive diarrhea. Cathartics are contraindicated in patients who have ingested corrosives and in those with preexisting diarrhea. Magnesium-containing cathartics should not be used in patients with renal failure.

Dilution (i.e., drinking water, another clear liquid, or milk at a volume of 5 mL/kg of body weight) is recommended only after the ingestion of corrosives (acids, alkali). It may increase the dissolution rate (and hence absorption) of capsules, tablets, and other solid ingestants and should *not* be used in these circumstances.

Endoscopic or surgical removal of poisons may be useful in rare situations, such as ingestion of a potentially toxic foreign body that fails to transit the gastrointestinal tract, a potentially lethal amount of a heavy metal (arsenic, iron, mercury, thallium), or agents that have coalesced into gastric concretions or bezoars (heavy metals, lithium, salicylates, sustained-release preparations). Patients who become toxic from cocaine due to its leakage from ingested drug packets require immediate surgical intervention.

Decontamination of Other Sites

Immediate, copious flushing with water, saline, or another available clear, drinkable liquid is the initial treatment for topical exposures (exceptions include alkali metals, calcium oxide, phosphorus). Saline is preferred for eye irrigation. A triple wash (water, soap, water) may be best for dermal decontamination. Inhalational exposures should be treated initially with fresh air or supplemental oxygen. The removal of liquids from body cavities such as the vagina or rectum is best accomplished by irrigation. Solids (drug packets, pills) should be removed manually, preferably under direct visualization.

ENHANCEMENT OF POISON ELIMINATION

Although the elimination of most poisons can be accelerated by therapeutic interventions, the pharmacokinetic efficacy (removal of drug at a rate greater than that accomplished by intrinsic elimination) and clinical benefit (shortened duration of toxicity or improved outcome) of such interventions are often more theoretical than proven. Accordingly, the decision to use such measures should be based on the actual or predicted toxicity and the potential efficacy, cost, and risks of therapy.

Multiple-Dose Activated Charcoal

Repetitive oral dosing with charcoal can enhance the elimination of previously absorbed substances by binding them within the gut as they are excreted in the bile, are secreted by gastrointestinal cells, or passively diffuse into the gut lumen (*reverse absorption* or *enterocapillary exsorption*). Doses of 0.5–1 g/kg of body weight every 2–4 h, adjusted downward to avoid regurgitation in patients with decreased gastrointestinal motility, are generally recommended. Pharmacokinetic efficacy approaches that of hemodialysis for some agents (e.g., phenobarbital, [theophylline](#)). Multiple-dose therapy should be considered only for selected agents ([theophylline](#), phenobarbital, carbamazepine, [dapsone](#), quinine). Complications include intestinal obstruction, pseudo-obstruction, and nonocclusive intestinal infarction in patients with decreased gut motility. Because of electrolyte and fluid shifts, sorbitol and other cathartics are absolutely contraindicated when multiple doses of activated charcoal are administered.

Urinary Alkalinization

Ion trapping via alteration of urine pH may prevent the renal reabsorption of poisons that undergo excretion by glomerular filtration and active tubular secretion. Since membranes are more permeable to non-ionized molecules than to their ionized counterparts, acidic (low- pK_a) poisons are ionized and trapped in alkaline urine, whereas basic ones become ionized and trapped in acid urine. Urinary alkalinization (producing a urine pH ≥ 7.5 and a urine output of 3–6 mL/kg of body weight per hour by the addition of [sodium bicarbonate](#) to an IV solution) enhances the excretion of chlorophenoxyacetic acid herbicides, chlorpropamide, diflunisal, fluoride, methotrexate, phenobarbital, sulfonamides, and salicylates. Contraindications include congestive heart failure, renal failure, and cerebral edema. Acid-base, fluid, and electrolyte parameters should be monitored carefully. Although acid diuresis may make theoretical sense for some overdoses (amphetamines), it is *never* indicated and is potentially harmful.

Extracorporeal Removal

Hemodialysis, charcoal or resin hemoperfusion, hemofiltration, plasmapheresis, and exchange transfusion are capable of removing any toxin from the bloodstream. Agents most amenable to enhanced elimination by dialysis have low molecular mass (<500 Da), high water solubility, low protein binding, small volumes of distribution (<1 L/kg of body weight), prolonged elimination (long half-life), and high dialysis clearance relative to total-body clearance. Molecular weight, water solubility, and protein binding do not limit the efficacy of the other forms of extracorporeal removal.

Dialysis should be considered in cases of severe poisoning due to carbamazepine, ethylene glycol, isopropyl alcohol, lithium, methanol, [theophylline](#), salicylates, and valproate. Although hemoperfusion may be more effective in removing some of these poisons, it does not correct associated acid-base and electrolyte abnormalities, and most hospitals no longer have hemoperfusion cartridges readily available. Fortunately, recent advances in hemodialysis technology make it as effective as hemoperfusion for removing poisons such as caffeine, carbamazepine, and [theophylline](#). Both techniques require central venous access and systemic anticoagulation and may result in transient hypotension. Hemoperfusion may also cause hemolysis, hypocalcemia, and thrombocytopenia. Peritoneal dialysis and exchange transfusion are less effective but may be used when other procedures are unavailable, contraindicated, or technically difficult (e.g., in infants). Exchange transfusion may be indicated in the treatment of severe arsine- or sodium chlorate-induced hemolysis, methemoglobinemia, and sulfhemoglobinemia. Although hemofiltration can enhance elimination of

aminoglycosides, vancomycin, and metal-chelate complexes, the roles of hemofiltration and plasmapheresis in the treatment of poisoning are not yet defined.

Candidates for extracorporeal removal therapies include patients with severe toxicity whose condition deteriorates despite aggressive supportive therapy; those with potentially prolonged, irreversible, or fatal toxicity; those with dangerous blood levels of toxins; those who lack the capacity for self-detoxification because of liver or renal failure; and those with a serious underlying illness or complication that will adversely affect recovery.

Other Techniques

The elimination of heavy metals can be enhanced by chelation, and the removal of carbon monoxide can be accelerated by hyperbaric oxygenation.

ADMINISTRATION OF ANTIDOTES

Antidotes counteract the effects of poisons by neutralizing them (e.g., antibody-antigen reactions, chelation, chemical binding) or by antagonizing their physiologic effects (e.g., activation of opposing nervous system activity, provision of a competitive metabolic or receptor substrate). Poisons or conditions with specific antidotes include acetaminophen, anticholinergic agents, anticoagulants, benzodiazepines, beta blockers, calcium channel blockers, carbon monoxide, cardiac glycosides, cholinergic agents, cyanide, drug-induced dystonic reactions, ethylene glycol, fluoride, heavy metals, hypoglycemic agents, [isoniazid](#), membrane-active agents, methemoglobinemia, opioids, sympathomimetics, and a variety of envenomations. Intravenous lipid emulsion has been shown to be a successful antidote for poisoning from various anesthetics and membrane-active agents (e.g., cyclic antidepressants), but the exact mechanism of benefit is still under investigation. Antidotes can significantly reduce morbidity and mortality rates but are potentially toxic if used for inappropriate reasons. Since their safe use requires correct identification of a specific poisoning or syndrome, details of antidotal therapy are discussed with the conditions for which they are indicated ([Table 450-4](#)).

PREVENTION OF REEXPOSURE

Poisoning is a preventable illness. Unfortunately, some adults and children are poison-prone, and recurrences are common. Unintentional polypharmacy poisoning has become especially common among adults with developmental delays, among the growing population of geriatric patients who are prescribed a large number of medications, and among adolescents and young adults experimenting with pharmaceuticals for recreational euphoria. Adults with unintentional exposures should be instructed regarding the safe use of medications and chemicals (according to labeling instructions). Confused patients may need assistance with the administration of their medications. Errors in dosing by health care providers may require educational efforts. Patients should be advised to avoid circumstances that result in chemical exposure or poisoning. Appropriate agencies and health departments (e.g., Occupational Health and Safety Administration [OSHA]) should be notified in cases of environmental or workplace exposure. The best approach to young children and patients with intentional overdose (deliberate self-harm or attempted suicide) is to limit their access to poisons. In households where children live or visit, alcoholic beverages, medications, household products (automotive, cleaning, fuel, pet-care, and toiletry products), inedible plants, and vitamins should be kept out of reach or in locked or child-proof cabinets. Depressed or psychotic patients should undergo psychiatric assessment, disposition, and follow-up.

They should be given prescriptions for a limited supply of drugs with a limited number of refills and should be monitored for compliance and response to therapy.

TABLE 450-3

Fundamentals of Poisoning Management

Supportive Care	
Airway protection	Treatment of seizures
Oxygenation/ventilation	Correction of temperature abnormalities
Treatment of arrhythmias	Correction of metabolic derangements
Hemodynamic support	Prevention of secondary complications
Prevention of Further Poison Absorption	
Gastrointestinal decontamination	Decontamination of other sites
Gastric lavage	Eye decontamination
Activated charcoal	Skin decontamination
Whole-bowel irrigation	Body cavity evacuation
Dilution	
Endoscopic/surgical removal	
Enhancement of Poison Elimination	
Multiple-dose activated charcoal administration Alteration of urinary pH Chelation	Extracorporeal removal Hemodialysis Hemoperfusion Hemofiltration Plasmapheresis Exchange transfusion Hyperbaric oxygenation
Administration of Antidotes	
Neutralization by antibodies	Metabolic antagonism
Neutralization by chemical binding	Physiologic antagonism

Prevention of Reexposure	
Adult education	Notification of regulatory agencies
Child-proofing	Psychiatric referral

TABLE 450-4

Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Stimulated				
Sympathetics ^a				

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Sympathomimetics	α_1 -Adrenergic agonists (decongestants): phenylephrine , phenylpropanolamine β_2 -Adrenergic agonists (bronchodilators): albuterol, terbutaline Nonspecific adrenergic agonists: amphetamines, cocaine, ephedrine	Stimulation of central and peripheral sympathetic receptors directly or indirectly (by promoting release or inhibiting reuptake of norepinephrine and sometimes dopamine)	Physiologic stimulation (Table 450-2). Reflex bradycardia can occur with selective α_1 agonists; β agonists can cause hypotension and hypokalemia.	Phentolamine, a nonselective α_1 -adrenergic receptor antagonist, for severe hypertension due to α_1 -adrenergic agonists; propranolol, a nonselective β blocker, for hypotension and tachycardia due to β_2 agonists; <i>either</i> labetalol , a β blocker with α -blocking activity, <i>or</i> phentolamine with esmolol, metoprolol , or another cardioselective β blocker for hypertension with tachycardia due to nonselective agents (β blockers, if used alone, can exacerbate hypertension and vasospasm due to unopposed α stimulation.); benzodiazepines; propofol

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Ergot alkaloids	Ergotamine, methysergide, bromocriptine, pergolide	Stimulation and inhibition of serotonergic and α - adrenergic receptors; stimulation of dopamine receptors	Physiologic stimulation (Table 450-2); formication; vasospasm with limb (isolated or generalized), myocardial, and cerebral ischemia progressing to gangrene or infarction. Hypotension, bradycardia, and involuntary movements can also occur.	Nitroprusside or nitroglycerine for severe vasospasm; prazosin (an α_1 blocker), captopril, nifedipine, and cyproheptadine (a serotonin receptor antagonist) for mild- to-moderate limb ischemia; dopamine receptor antagonists (antipsychotics) for hallucinations and movement disorders

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Methylxanthines	Caffeine, theophylline	Inhibition of adenosine synthesis and adenosine receptor antagonism; stimulation of epinephrine and norepinephrine release; inhibition of phosphodiesterase resulting in increased intracellular cyclic adenosine and guanosine monophosphate	Physiologic stimulation (Table 450-2); pronounced gastrointestinal symptoms and β agonist effects (see above). Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Propranolol, a nonselective β blocker, for tachycardia with hypotension; any β blocker for supraventricular or ventricular tachycardia without hypotension; elimination enhanced by multiple-dose charcoal, hemoperfusion, and hemodialysis. Indications for hemoperfusion or hemodialysis include unstable vital signs, seizures, and a theophylline level of 80–100 $\mu\text{g/mL}$ after an acute overdose and 40–60 $\mu\text{g/mL}$ with chronic exposure.

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Monoamine oxidase inhibitors	Phenelzine, tranylcypromine, selegiline	Inhibition of monoamine oxidase resulting in impaired metabolism of endogenous catecholamines and exogenous sympathomimetic agents	Delayed or slowly progressive physiologic stimulation (Table 450-2); terminal hypotension and bradycardia in severe cases	Short-acting agents (e.g., nitroprusside, esmolol) for severe hypertension and tachycardia; direct- acting sympathomimetics (e.g., norepinephrine, epinephrine) for hypotension and bradycardia
Anticholinergics				
Antihistamines	Diphenhydramine , doxylamine, pyrilamine	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors. At high doses, amantadine, diphenhydramine , orphenadrine, phenothiazines, and tricyclic antidepressants have additional nonanticholinergic activity (see below).	Physiologic stimulation (Table 450-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and non-anticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Antipsychotics	Chlorpromazine, olanzapine, quetiapine, thioridazine	Inhibition of α -adrenergic, dopaminergic, histaminergic, muscarinic, and serotonergic receptors. Some agents also inhibit sodium, potassium, and calcium channels.	Physiologic depression (Table 450-2), miosis, anticholinergic effects (see above), extrapyramidal reactions (see below), tachycardia	Sodium bicarbonate for ventricular tachydysrhythmias associated with QRS prolongation; magnesium, isoproterenol, and overdrive pacing for torsades des pointes. Avoid class IA, IC, and III antiarrhythmics.
Belladonna alkaloids	Atropine , hyoscyamine, scopolamine	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors	Physiologic stimulation (Table 450-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and non-anticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Cyclic antidepressants	Amitriptyline, doxepin , imipramine	Inhibition of α -adrenergic, dopaminergic, GABA-ergic, histaminergic, muscarinic, and serotonergic receptors; inhibition of sodium channels (see membrane-active agents); inhibition of norepinephrine and serotonin reuptake	Physiologic depression (Table 450-2), seizures, tachycardia, cardiac conduction delays (increased PR, QRS, JT, and QT intervals; terminal QRS right-axis deviation) with aberrancy and ventricular tachydysrhythmias; anticholinergic toxidrome (see above)	Hypertonic sodium bicarbonate (or hypertonic saline) for ventricular tachydysrhythmias associated with QRS prolongation. Use of phenytoin is controversial. Avoid class IA, IC, and III antiarrhythmics. IV emulsion therapy may be beneficial in some cases.
Mushrooms and plants	<i>Amanita muscaria</i> and <i>A. pantherina</i> , henbane, jimson weed, nightshade	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors	Physiologic stimulation (Table 450-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and nonanticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).
Depressed				
Sympatholytics				

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
α_2 -Adrenergic agonists	Clonidine, guanabenz, tetrahydrozoline and other imidazoline decongestants, tizanidine and other imidazoline muscle relaxants	Stimulation of α_2 - adrenergic receptors leading to inhibition of CNS sympathetic outflow. Activity at nonadrenergic imidazoline binding sites also contributes to CNS effects.	Physiologic depression (Table 450-2), miosis. Transient initial hypertension may be seen.	Dopamine and norepinephrine for hypotension; atropine for symptomatic bradycardia; naloxone for CNS depression (inconsistently effective)
Antipsychotics	Chlorpromazine, clozapine, haloperidol, risperidone, thioridazine	Inhibition of α - adrenergic, dopaminergic, histaminergic, muscarinic, and serotonergic receptors. Some agents also inhibit sodium, potassium, and calcium channels.	Physiologic depression (Table 450-2), miosis, anticholinergic effects (see above), extrapyramidal reactions (see below), tachycardia. Cardiac conduction delays (increased PR, QRS, JT, and QT intervals) with ventricular tachydysrhythmias, including torsades des pointes, can sometimes develop.	Sodium bicarbonate for ventricular tachydysrhythmias associated with QRS prolongation; magnesium, isoproterenol, and overdrive pacing for torsades des pointes. Avoid class IA, IC, and III antiarrhythmics.

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
β -Adrenergic blockers	Cardioselective (β_1) blockers: atenolol, esmolol, metoprolol Nonselective (β_1 and β_2) blockers: nadolol, propranolol, timolol Partial β agonists: acebutolol, pindolol α_1 Antagonists: carvedilol, labetalol Membrane-active agents: acebutolol, propranolol, sotalol	Inhibition of β - adrenergic receptors (class II antiarrhythmic effect). Some agents have activity at additional receptors or have membrane effects (see below).	Physiologic depression (Table 450-2), atrioventricular block, hypoglycemia, hyperkalemia, seizures. Partial agonists can cause hypertension and tachycardia. Sotalol can cause increased QT interval and ventricular tachydysrhythmias. Onset may be delayed after sotalol and sustained- release formulation overdose.	Glucagon for hypotension and symptomatic bradycardia. Atropine , isoproterenol, dopamine, dobutamine, epinephrine , and norepinephrine may sometimes be effective. High-dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), electrical pacing, and mechanical cardiovascular support for refractory cases.

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Calcium channel blockers	Diltiazem, nifedipine and other dihydropyridine derivatives, verapamil	Inhibition of slow (type L) cardiovascular calcium channels (class IV antiarrhythmic effect)	<p>Physiologic depression (Table 450-2), atrioventricular block, organ ischemia and infarction, hyperglycemia, seizures.</p> <p>Hypotension is usually due to decreased vascular resistance rather than to decreased cardiac output.</p> <p>Onset may be delayed for ≥ 12 h after overdose of sustained-release formulations.</p>	<p>Calcium and glucagon for hypotension and symptomatic bradycardia.</p> <p>Dopamine, epinephrine, norepinephrine, atropine, and isoproterenol are less often effective but can be used adjunctively. High- dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), IV lipid emulsion therapy, electrical pacing, and mechanical cardiovascular support for refractory cases.</p>

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Cardiac glycosides	Digoxin, endogenous cardioactive steroids, foxglove and other plants, toad skin secretions (Bufonidae spp.)	Inhibition of cardiac Na^+ , K^+ -ATPase membrane pump	Physiologic depression (Table 450-2); gastrointestinal, psychiatric, and visual symptoms; atrioventricular block with or without concomitant supraventricular tachyarrhythmia; ventricular tachyarrhythmias; hyperkalemia in acute poisoning. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Digoxin-specific antibody fragments for hemodynamically compromising dysrhythmias, Mobitz II or third- degree atrioventricular block, hyperkalemia (>5.5 meq/L; in acute poisoning only). Temporizing measures include atropine , dopamine, epinephrine , and external cardiac pacing for bradydysrhythmias and magnesium, lidocaine , or phenytoin, for ventricular tachydysrhythmias. Internal cardiac pacing and cardioversion can increase ventricular irritability and should be reserved for refractory cases.

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Cyclic antidepressants	Amitriptyline, doxepin , imipramine	Inhibition of α -adrenergic, dopaminergic, GABA-ergic, histaminergic, muscarinic, and serotonergic receptors; inhibition of sodium channels (see membrane-active agents); inhibition of norepinephrine and serotonin reuptake	Physiologic depression (Table 450-2), seizures, tachycardia, cardiac conduction delays (increased PR, QRS, JT, and QT intervals; terminal QRS right-axis deviation) with aberrancy and ventricular tachydysrhythmias; anticholinergic toxidrome (see above)	Hypertonic sodium bicarbonate (or hypertonic saline) for ventricular tachydysrhythmias associated with QRS prolongation. Use of phenytoin is controversial. Avoid class IA, IC, and III antiarrhythmics. IV emulsion therapy may be beneficial in some cases.
Cholinergics				
Acetylcholinesterase inhibitors	Carbamate insecticides (aldicarb, carbaryl, propoxur) and medicinals (neostigmine, physostigmine, tacrine); nerve gases (sarin, soman, tabun, VX); organophosphate insecticides (diazinon, chlorpyrifos-ethyl, malathion)	Inhibition of acetylcholinesterase leading to increased synaptic acetylcholine at muscarinic and nicotinic cholinergic receptor sites	Physiologic depression (Table 450-2). Muscarinic signs and symptoms: seizures, excessive secretions (lacrimation, salivation, bronchorrhea and wheezing, diaphoresis), and increased bowel and bladder activity with nausea, vomiting, diarrhea, abdominal cramps, and incontinence of feces and urine. Nicotinic signs and symptoms: hypertension,	Atropine for muscarinic signs and symptoms; 2-PAM, a cholinesterase reactivator, for nicotinic signs and symptoms due to organophosphates, nerve gases, or an unknown anticholinesterase
Muscarinic agonists	Bethanechol , mushrooms (<i>Boletus</i> , <i>Clitocybe</i> , <i>Inocybe</i> spp.), pilocarpine	Stimulation of CNS and postganglionic parasympathetic cholinergic (muscarinic) receptors		

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Nicotinic agonists	Lobeline, nicotine (tobacco)	Stimulation of preganglionic sympathetic and parasympathetic and striated muscle (neuromuscular junction) cholinergic (nicotine) receptors	tachycardia, muscle cramps, fasciculations, weakness, and paralysis. Death is usually due to respiratory failure. Cholinesterase activity in plasma and red cells is <50% of normal in acetylcholinesterase inhibitor poisoning.	
Sedative-hypnotics ^b				
Anticonvulsants	Carbamazepine, ethosuximide, felbamate, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenytoin, tiagabine, topiramate, valproate, zonisamide	Potential of the inhibitory effects of GABA by binding to the neuronal GABA-A chloride channel receptor complex and increasing the frequency or duration of chloride channel opening in response to GABA stimulation. Baclofen and, to some extent, GHB act at the GABA-B receptor complex. Meprobamate, its metabolite carisoprodol, felbamate, and orphenadrine antagonize NMDA excitatory receptors. Ethosuximide,	Physiologic depression (Table 450-2), nystagmus. Delayed absorption can occur with carbamazepine, phenytoin, and valproate. Myoclonus, seizures, hypertension, and tachyarrhythmias can occur with baclofen, carbamazepine, and orphenadrine.	Benzodiazepines, barbiturates, or propofol for seizures.
Barbiturates	Short-acting: butabarbital, pentobarbital, secobarbital Long-acting: phenobarbital, primidone			Hemodialysis and hemoperfusion may be indicated for severe poisoning by some agents (see “Extracorporeal Removal,” in text).

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Benzodiazepines	Ultrashort-acting: estazolam, midazolam, temazepam, triazolam Short-acting: alprazolam, flunitrazepam, lorazepam , oxazepam Long-acting: chlordiazepoxide, clonazepam, diazepam, flurazepam Pharmacologically related agents: zaleplon, zolpidem	valproate, and zonisamide decrease conduction through T-type calcium channels. Valproate decreases GABA degradation, and tiagabine blocks GABA reuptake. Carbamazepine, lamotrigine , oxcarbazepine , phenytoin, topiramate, valproate, and zonisamide slow the rate of recovery of	Tachyarrhythmias can also occur with chloral hydrate. AGMA, hypernatremia, hyperosmolality, hyperammonemia, chemical hepatitis, and hypoglycemia can be seen in valproate poisoning. Carbamazepine and oxcarbazepine may produce hyponatremia from SIADH.	See above and below for treatment of anticholinergic and sodium channel (membrane)– blocking effects.
GABA precursors	γ-Hydroxybutyrate (sodium oxybate; GHB), γ-butyrolactone (GBL), 1,4-butanediol	inactivated sodium channels. Some agents also have α ₂ agonist, anticholinergic, and sodium channel– blocking activity (see above and below).	Some agents can cause anticholinergic and sodium channel (membrane) blocking effects (see above and below).	
Muscle relaxants	Baclofen , carisoprodol, cyclobenzaprine , etomidate, metaxalone, methocarbamol , orphenadrine, propofol, tizanidine and other imidazoline muscle relaxants	Baclofen acts at GABA-B receptor complex; Stimulation of α ₂ -adrenergic receptors inhibits CNS sympathetic outflow. Activity at nonadrenergic imidazoline binding sites also contributes to CNS effects. The others have centrally-acting and	Physiologic depression (Table 450-2)	Goal-directed supportive care; benzodiazepines and barbiturates for seizures

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	various other MECHANISM OF ACTION mechanisms of	CLINICAL FEATURES	SPECIFIC TREATMENTS
Other agents	Chloral hydrate, ethchlorvynol, glutethimide, meprobamate, methaqualone, methypylon	action		
Discordant				
Asphyxiants				

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Cytochrome oxidase inhibitors	Cyanide, hydrogen sulfide	Inhibition of mitochondrial cytochrome oxidase, with consequent blockage of electron transport and oxidative metabolism. Carbon monoxide also binds to hemoglobin and myoglobin and prevents oxygen binding, transport, and tissue uptake. (Binding to hemoglobin shifts the oxygen dissociation curve to the left.)	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table 450-2); lactic acidosis; normal PO_2 and calculated oxygen saturation but decreased oxygen saturation by co-oximetry. (That measured by pulse oximetry is falsely elevated but is less than normal and less than the calculated value.) Headache and nausea are common with carbon monoxide. Sudden collapse may occur with cyanide and hydrogen sulfide exposure. A bitter almond breath odor may be noted with cyanide ingestion, and hydrogen sulfide smells like rotten eggs.	High-dose oxygen; IV hydroxocobalamin or IV sodium nitrite and sodium thiosulfate (Lilly cyanide antidote kit) for coma, metabolic acidosis, and cardiovascular dysfunction in cyanide poisoning or victims from a fire

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Methemoglobin inducers	Aniline derivatives, dapson , local anesthetics, nitrates, nitrites, nitrogen oxides, nitro- and nitrosohydrocarbons, phenazopyridine , primaquine-type antimalarials, sulfonamides	Oxidation of hemoglobin iron from ferrous (Fe^{2+}) to ferric (Fe^{3+}) state prevents oxygen binding, transport, and tissue uptake. (Methemoglobinemia shifts oxygen dissociation curve to the left.) Oxidation of hemoglobin protein causes hemoglobin precipitation and hemolytic anemia (manifesting as Heinz bodies and “bite cells” on peripheral- blood smear).	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table 450-2), gray-brown cyanosis unresponsive to oxygen at methemoglobin fractions >15–20%, headache, lactic acidosis (at methemoglobin fractions >45%), normal Po_2 and calculated oxygen saturation but decreased oxygen saturation and increased methemoglobin fraction by co- oximetry (Oxygen saturation by pulse oximetry may be falsely increased or decreased but is less than normal and less than the calculated value.)	High-dose oxygen; IV methylene blue for methemoglobin fraction >30%, symptomatic hypoxemia, or ischemia (contraindicated in G6PD deficiency); exchange transfusion and hyperbaric oxygen for severe or refractory cases

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
AGMA inducers	Ethylene glycol	Ethylene glycol causes CNS depression and increased serum osmolality. Metabolites (primarily glycolic acid) cause AGMA, CNS depression, and renal failure. Precipitation of oxalic acid metabolite as calcium salt in tissues and urine results in hypocalcemia, tissue edema, and crystalluria.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap, calcium oxylate crystalluria; delayed AGMA, back pain, renal failure; coma, seizures, hypotension, ARDS in severe cases	Sodium bicarbonate to correct acidemia; thiamine , folic acid, magnesium, and high-dose pyridoxine to facilitate metabolism; ethanol or fomepizole for AGMA, crystalluria or renal dysfunction, ethylene glycol level >3 mmol/L (20 mg/dL), and ethanol-like intoxication or increased osmolal gap if level not readily obtainable; hemodialysis for persistent AGMA, lack of clinical improvement, and renal dysfunction; hemodialysis also useful for enhancing ethylene glycol elimination and shortening duration of treatment when ethylene glycol level is >8 mmol/L (50 mg/dL).

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
AGMA inducers	Iron	Hydration of ferric (Fe^{3+}) ion generates H^+ . Non-transferrin-bound iron catalyzes formation of free radicals that cause mitochondrial injury, lipid peroxidation, increased capillary permeability, vasodilation, and organ toxicity.	Initial nausea, vomiting, abdominal pain, diarrhea; AGMA, cardiovascular and CNS depression, hepatitis, coagulopathy, and seizures in severe cases. Radiopaque iron tablets may be seen on abdominal x-ray.	Whole-bowel irrigation for large ingestions; endoscopy and gastrostomy if clinical toxicity and large number of tablets are still visible on x-ray; IV hydration; sodium bicarbonate for acidemia; IV deferoxamine for systemic toxicity, iron level $>90 \mu\text{mol/L}$ ($500 \mu\text{g/dL}$)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
	Methanol	Methanol causes ethanol-like CNS depression and increased serum osmolality. Formic acid metabolite causes AGMA and retinal toxicity.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap; delayed AGMA, visual (clouding, spots, blindness) and retinal (edema, hyperemia) abnormalities; coma, seizures, cardiovascular depression in severe cases; possible pancreatitis	Gastric aspiration for recent ingestion; sodium bicarbonate to correct acidemia; high-dose folinic acid or folate to facilitate metabolism; ethanol or fomepizole for AGMA, visual symptoms, methanol level >6 mmol/L (20 mg/dL), and ethanol-like intoxication or increased osmolal gap if level not readily obtainable; hemodialysis for persistent AGMA, lack of clinical improvement, and renal dysfunction; hemodialysis also useful for enhancing methanol elimination and shortening duration of treatment when methanol level is >15 mmol/L (50 mg/dL)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
	Salicylate	Increased sensitivity of CNS respiratory center to changes in and stimulates respiration. Uncoupling of oxidative phosphorylation, inhibition of Krebs cycle enzymes, and stimulation of carbohydrate and lipid metabolism generate unmeasured endogenous anions and cause AGMA.	Initial nausea, vomiting, hyperventilation, alkalemia, alkaluria; subsequent alkalemia with both respiratory alkalosis and AGMA and paradoxical aciduria; late acidemia with CNS and respiratory depression; cerebral and pulmonary edema in severe cases. Hypoglycemia, hypocalcemia, hypokalemia, and seizures can occur.	IV hydration and supplemental glucose; sodium bicarbonate to correct acidemia; urinary alkalinization for systemic toxicity; hemodialysis for coma, cerebral edema, seizures, pulmonary edema, renal failure, progressive acid-base disturbances or clinical toxicity, salicylate level >7 mmol/L (100 mg/dL) following acute overdose
CNS syndromes				
Extrapyramidal reactions	Antipsychotics (see above), some cyclic antidepressants and antihistamines	Decreased CNS dopaminergic activity with relative excess of cholinergic activity	Akathisia, dystonia, parkinsonism	Oral or parenteral anticholinergic agent such as benztropine or diphenhydramine

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Isoniazid		Interference with activation and supply of pyridoxal-5-phosphate, a cofactor for glutamic acid decarboxylase, which converts glutamic acid to GABA, results in decreased levels of this inhibitory CNS neurotransmitter; complexation with and depletion of pyridoxine itself; inhibition of nicotine adenine dinucleotide-dependent lactate and hydroxybutyrate dehydrogenases, resulting in substrate accumulation	Nausea, vomiting, agitation, confusion; coma, respiratory depression, seizures, lactic and ketoacidosis in severe cases	High-dose IV pyridoxine (vitamin B ₆) for agitation, confusion, coma, and seizures; diazepam or barbiturates for seizures

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Lithium		Interference with cell membrane ion transport, adenylate cyclase and Na^+ , K^+ -ATPase activity, and neurotransmitter release	Nausea, vomiting, diarrhea, ataxia, choreoathetosis, encephalopathy, hyperreflexia, myoclonus, nystagmus, nephrogenic diabetes insipidus, falsely elevated serum chloride with low anion gap, tachycardia; coma, seizures, arrhythmias, hyperthermia, and prolonged or permanent encephalopathy and movement disorders in severe cases; delayed onset after acute overdose, particularly with delayed-release formulations. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Whole-bowel irrigation for large ingestions; IV hydration; hemodialysis for coma, seizures, encephalopathy or neuromuscular dysfunction (severe, progressive, or persistent), peak lithium level >4 meq/L following acute overdose

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Serotonin syndrome	Amphetamines, cocaine, dextromethorphan, meperidine, MAO inhibitors, selective serotonin (5-HT) reuptake inhibitors, tricyclic antidepressants, tramadol, triptans, tryptophan	Promotion of serotonin release, inhibition of serotonin reuptake, or direct stimulation of CNS and peripheral serotonin receptors (primarily 5-HT-1a and 5-HT-2), alone or in combination	Altered mental status (agitation, confusion, mutism, coma, seizures), neuromuscular hyperactivity (hyperreflexia, myoclonus, rigidity, tremors), and autonomic dysfunction (abdominal pain, diarrhea, diaphoresis, fever, flushing, labile hypertension, mydriasis, tearing, salivation, tachycardia). Complications include hyperthermia, lactic acidosis, rhabdomyolysis, and multisystem organ failure.	Discontinue the offending agent(s); the serotonin receptor antagonist cyproheptadine may be helpful in severe cases.
Membrane-active agent	Amantadine, antiarrhythmics (class I and III agents; some β blockers), antipsychotics (see above), antihistamines (particularly diphenhydramine), carbamazepine, local anesthetics (including	Blockade of fast sodium membrane channels prolongs phase 0 (depolarization) of the cardiac action potential, which prolongs QRS duration and promotes reentrant (monomorphic)	QRS and JT prolongation (or both) with hypotension, ventricular tachyarrhythmias, CNS depression, seizures; anticholinergic effects with amantadine,	Hypertonic sodium bicarbonate (or hypertonic saline) for cardiac conduction delays and monomorphic ventricular tachycardia; lidocaine for monomorphic ventricular

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
	cocaine), opioids (meperidine, propoxyphene), orphenadrine, quinoline antimalarials (chloroquine, hydroxychloroquine, quinine), cyclic antidepressants (see above)	ventricular tachycardia. Class Ia, Ic, and III antiarrhythmics also block potassium channels during phases 2 and 3 (repolarization) of the action potential, prolonging the JT interval and promoting early after-depolarizations and polymorphic (torsades des pointes) ventricular tachycardia. Similar effects on neuronal membrane channels cause CNS dysfunction. Some agents also block α -adrenergic and cholinergic receptors or have opioid effects (see above and Chap. 446).	antihistamines, carbamazepine, disopyramide, antipsychotics, and cyclic antidepressants (see above); opioid effects with meperidine and propoxyphene (see Chap. 446); cinchonism (hearing loss, tinnitus, nausea, vomiting, vertigo, ataxia, headache, flushing, diaphoresis), and blindness with quinoline antimalarials	tachycardia (except when due to class Ib antiarrhythmics); magnesium, isoproterenol, and overdrive pacing for polymorphic ventricular tachycardia; physostigmine for anticholinergic effects (see above); naloxone for opioid effects (see Chap. 446); extracorporeal removal for some agents (see text).

a

See above and [Chap. 447](#). ^bSee above and [Chap. 446](#).

Abbreviations: AGMA, anion-gap metabolic acidosis; ARDS, adult respiratory distress syndrome; CNS, central nervous system; GABA, γ -aminobutyric acid; GBL, γ -butyrolactone; GHB, γ -hydroxybutyrate; G6PD, glucose-6-phosphate dehydrogenase; MAO, monoamine oxidase; NDMA, *N*-methyl-D-aspartate; 2-PAM, pralidoxime; SIADH, syndrome of inappropriate antidiuretic hormone secretion.

SPECIFIC TOXIC SYNDROMES AND POISONINGS

Table 450-4 summarizes the pathophysiology, clinical features, and treatment of toxidromes and poisonings that are common, produce life-threatening toxicity, or require unique therapeutic interventions. In all cases, treatment should include attention to the general principles discussed above and, in particular, supportive care. Poisonings not covered in this chapter are discussed in specialized texts.

Alcohol, cocaine, hallucinogen, and opioid poisoning and alcohol and opioid withdrawal are discussed in Chaps. 445–447; nicotine addiction is discussed in Chap. 448; acetaminophen poisoning is discussed in Chap. 333; the neuroleptic malignant syndrome is discussed in Chap. 427; and heavy metal poisoning is discussed in Chap. 449.

GLOBAL CONSIDERATIONS



Risks of poisoning in the United States and throughout the world are in transition. Patterns of travel, immigration, and internet consumerism should always be considered in patients suspected of poisoning or overdose without a clear etiology. Immigrants into various countries may have underlying poisoning from various metals from work or the environment where they previously lived; herbal remedies, food products, and cosmetics imported from overseas may be contaminated with metals, toxic plants, or other pharmaceutical contaminants; and new drugs of abuse that originate in one part of the world quickly circulate due to the ease afforded by the internet. Expanding the history at the time of evaluation, recruiting the assistance of global health specialists, and ordering expanded laboratory panels may be indicated.

FURTHER READING

Dart RC et al: Expert consensus guidelines for stocking of antidotes in hospitals that provide emergency care. *Ann Emerg Med* 54:386, 2009. [[PubMed: 19406507](#)]

Hoffman RS et al (eds): *Goldfrank's Toxicologic Emergencies*, 10th ed. New York, McGraw-Hill, 2015.

Mazer-Amirshahi M et al: Trends in Emergency Department Resource Utilization for Poisoning-Related Visits, 2003–2011. *J Med Toxicol* 12:248, 2016. [[PubMed: 27342464](#)]

Mowry JB et al: 2015 Annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 33rd Annual Report. *Clin Toxicol* 54:924, 2016.

Thompson TM et al: The general approach to the poisoned patient. *Dis Mon* 60:509, 2014. [[PubMed: 25454682](#)]

Welker K, Mycyk MB: Pharmacology in the geriatric patient. *Emerg Med Clin North Am* 34:469, 2016. [[PubMed: 27475010](#)]

McGraw Hill

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is 178.250.250.21

Chapter A1: Atlas of Rashes Associated with Fever

Kenneth M. Kaye; Elaine T. Kaye

Given the extremely broad differential diagnosis, the presentation of a patient with fever and rash often poses a thorny diagnostic challenge for even the most astute and experienced clinician. Rapid narrowing of the differential by prompt recognition of a rash's key features can result in appropriate and sometimes life-saving therapy. This atlas presents high-quality images of a variety of rashes that are associated with fever, most of which have an infectious etiology.

FIGURE A1-1

A. Erythema leading to “slapped cheeks” appearance in erythema infectiosum (fifth disease) caused by parvovirus B19. **B.** Lacy reticular rash of erythema infectiosum. (*Panel A reprinted from K Wolff, RA Johnson: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



A

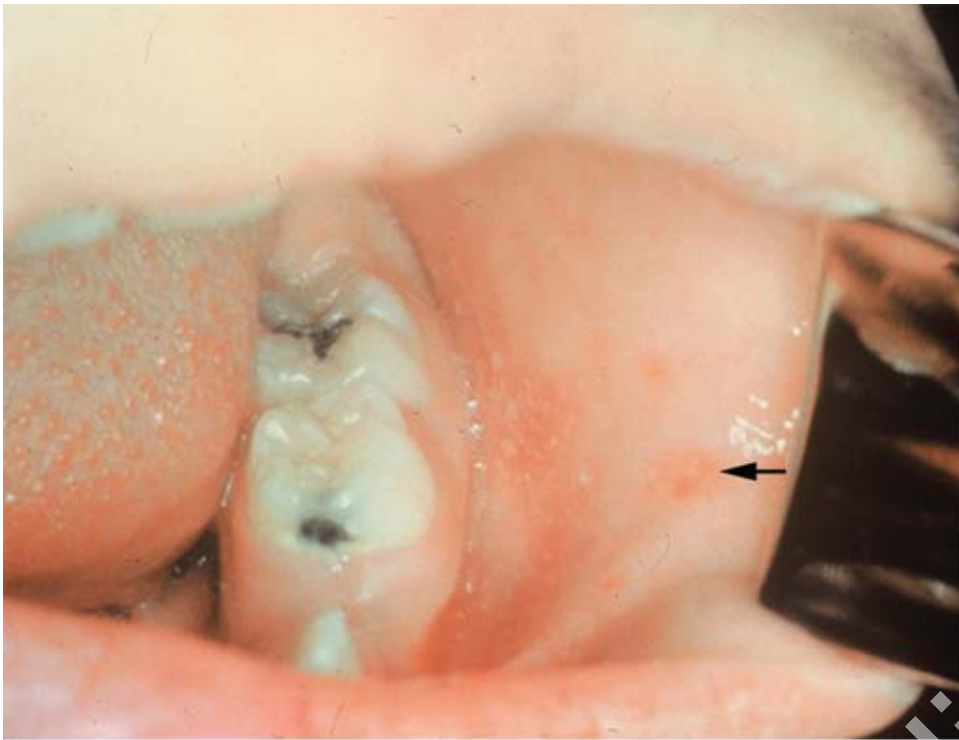
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-2

Koplik's spots, which manifest as white or bluish lesions with an erythematous halo on the buccal mucosa, usually occur in the first 2 days of measles symptoms and may briefly overlap the measles exanthem. The presence of the erythematous halo (*arrow indicates one example*) differentiates Koplik's spots from Fordyce's spots (ectopic sebaceous glands), which occur in the mouths of healthy individuals. (*Courtesy of the Centers for Disease Control and Prevention.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-3

In measles, discrete erythematous lesions become confluent on the face and neck over 2–3 days as the rash spreads downward to the trunk and arms, where lesions remain discrete. (*Reprinted from K Wolff, RA Johnson: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-4

In rubella, an erythematous exanthem spreads from the hairline downward and clears as it spreads. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-5

Exanthem subitum (roseola), caused by human herpesvirus 6, occurs most commonly in young children. A diffuse maculopapular exanthem follows resolution of fever. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-6

Erythematous macules and papules are apparent on the trunk and arm of this patient with acute HIV infection.

(Reprinted from K Wolff, RA Johnson: *Color Atlas and Synopsis of Clinical Dermatology*, 5th ed. New York, McGraw-Hill, 2005.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-7

This exanthematous, drug-induced eruption consists of brightly erythematous macules and papules, some of which are confluent, distributed symmetrically on the trunk and extremities. Ampicillin caused this rash. (Reprinted from K Wolff, RA Johnson: *Color Atlas and Synopsis of Clinical Dermatology*, 5th ed. New York, McGraw-Hill, 2005.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-8

Erythema migrans is the early cutaneous manifestation of Lyme disease and is characterized by an erythematous patch, which may be confluent or annular and sometimes has a target appearance. (*Reprinted from RP Usatine et al: Color Atlas of Family Medicine, 2nd ed. New York, McGraw-Hill, 2013. Courtesy of Thomas Corson, MD.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-9

Rose spots are evident as erythematous macules on the trunk of this patient with typhoid fever. (Courtesy of the Centers for Disease Control and Prevention.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-10

Systemic lupus erythematosus showing prominent malar erythema and minimal scaling. Involvement of other sun-exposed sites is also common. (Reprinted from K Wolff, RA Johnson: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 6th ed. New York, McGraw-Hill, 2009.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-11

Subacute lupus erythematosus on the upper chest, with brightly erythematous and slightly edematous coalescent papules and plaques. (*Reprinted from K Wolff, RA Johnson: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-12

Chronic discoid lupus erythematosus. Violaceous, hyperpigmented, atrophic plaques, often with evidence of follicular plugging (which may result in scarring), are characteristic of this cutaneous form of lupus. (*Reprinted from K Wolff, RA Johnson, AP Saavedra: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 7th ed. New York, McGraw-Hill, 2013.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine. 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-13

The rash of Still's disease typically exhibits evanescent, erythematous papules that appear at the height of fever on the trunk and proximal extremities. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-14

Impetigo is a superficial group A streptococcal or *Staphylococcus aureus* infection consisting of honey-colored crusts and erythematous weeping erosions. (*Reprinted from K Wolff, RA Johnson: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-15

Erysipelas is a group A streptococcal infection of the superficial dermis and consists of well-demarcated, erythematous, edematous, warm plaques. (*Reprinted from K Wolff, RA Johnson, AP Saavedra: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 7th ed. New York, McGraw-Hill, 2013.*)

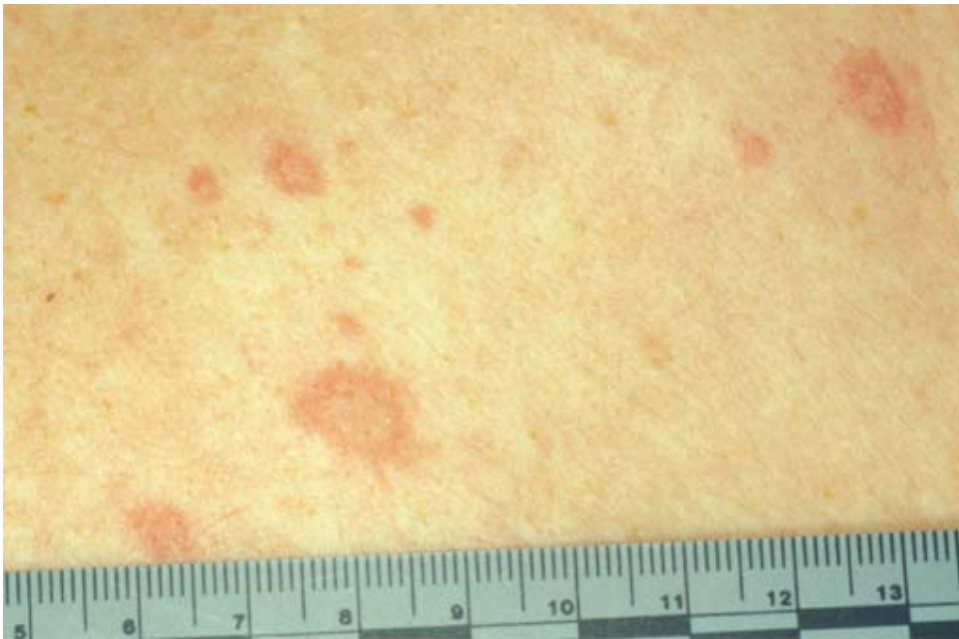


Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-16

Top: Petechial lesions of Rocky Mountain spotted fever on the lower legs and soles of a young, otherwise healthy patient. **Bottom:** Close-up of lesions from the same patient. (Courtesy of Lindsey Baden, MD; with permission.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-17

Primary syphilis with firm, nontender chancres. (*Courtesy of M. Rein and the Centers for Disease Control and Prevention.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-18

Secondary syphilis, demonstrating a papulosquamous truncal eruption.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-19

Secondary syphilis commonly affects the palms and soles with scaling, firm, red-brown papules.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-20

Condylomata lata are moist, somewhat verrucous intertriginous plaques seen in secondary syphilis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-21

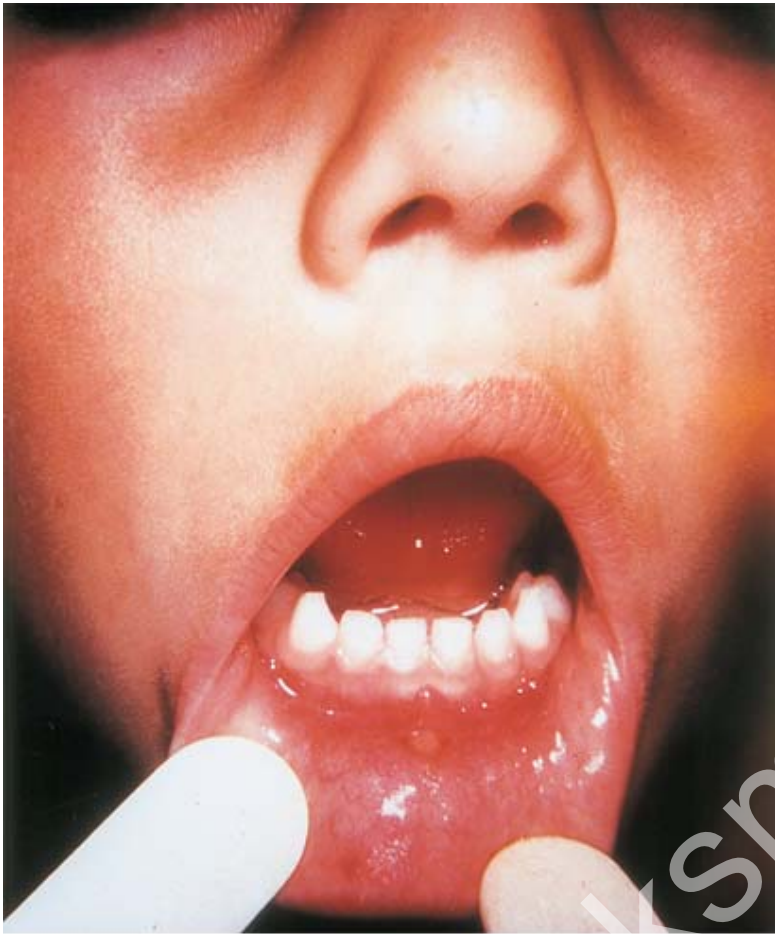
Mucous patches on the tongue of a patient with secondary syphilis. (*Courtesy of Ron Roddy; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-22

Tender vesicles and erosions in the mouth of a patient with hand-foot-and-mouth disease. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-23

Septic emboli with hemorrhage and infarction due to acute *Staphylococcus aureus* endocarditis. (*Courtesy of Lindsey Baden, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-24

Erythema multiforme is characterized by erythematous plaques with a target or iris morphology, sometimes with a vesicle in the center. It usually results from a hypersensitivity reaction to infections (especially with herpes simplex virus or *Mycoplasma pneumoniae*) or drugs. (Reprinted from K Wolff, RA Johnson: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 6th ed. New York, McGraw-Hill, 2009.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-25

Scarlet fever exanthem. Finely punctate erythema has become confluent (scarlatiniform); accentuation of linear erythema in body folds (Pastia's lines) is seen here. (*Reprinted from K Wolff, RA Johnson: Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

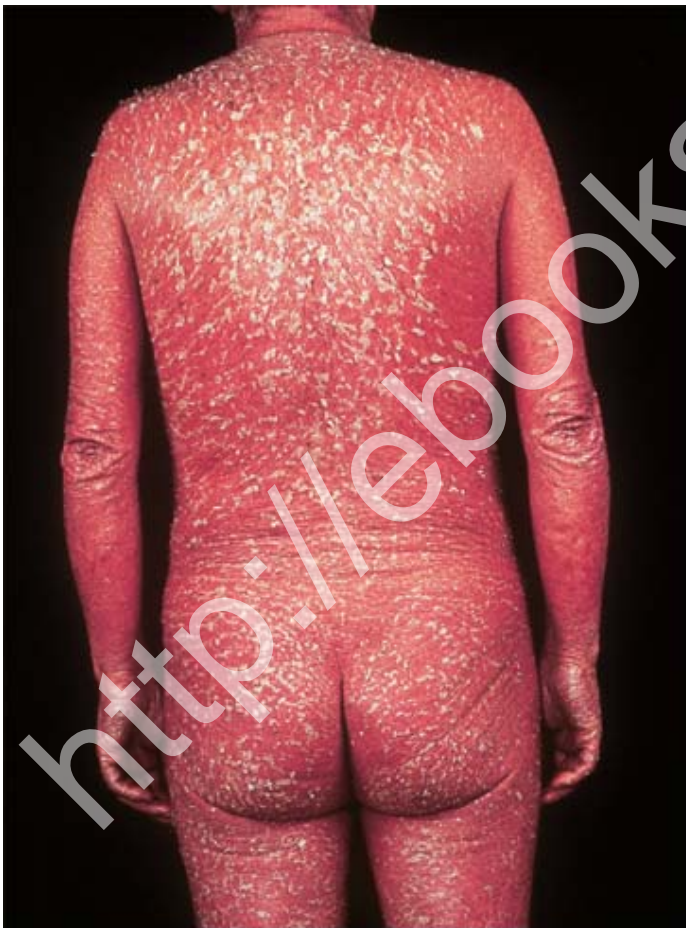
FIGURE A1-26

Erythema progressing to bullae with resulting sloughing of the entire thickness of the epidermis occurs in toxic epidermal necrolysis. This reaction was due to a sulfonamide. (*Reprinted from K Wolff, RA Johnson: Color Atlas and Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-27

Diffuse erythema and scaling are present in this patient with psoriasis and the exfoliative erythroderma syndrome.
(Reprinted from K Wolff, RA Johnson: *Color Atlas and Synopsis of Clinical Dermatology*, 6th ed. New York, McGraw-Hill, 2009.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-28

This infant with **staphylococcal scalded skin syndrome** demonstrates generalized desquamation. (*Reprinted from K Wolff, RA Johnson: Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-29

Fissuring of the lips and an erythematous exanthem are evident in this patient with Kawasaki disease. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-30

Numerous varicella lesions at various stages of evolution: vesicles on an erythematous base and umbilicated vesicles, which then develop into crusting lesions. (*Courtesy of the Centers for Disease Control and Prevention.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-31

Lesions of disseminated zoster at different stages of evolution, including pustules and crusted lesions similar to varicella. Note nongrouping of lesions, in contrast to herpes simplex or zoster (shingles). (*Reprinted from K Wolff, RA Johnson, AP Saavedra: Color Atlas and Synopsis of Clinical Dermatology, 7th ed. New York, McGraw-Hill, 2013.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-32

Herpes zoster is seen in this patient taking **prednisone**. Grouped vesicles and crusted lesions are seen in the T2 dermatome on the back and arm (**A**) and on the right side of the chest (**B**). (Reprinted from K Wolff, RA Johnson: *Color Atlas and Synopsis of Clinical Dermatology*, 6th ed. New York, McGraw-Hill, 2009.)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-33

A. Eschar at the site of the mite bite in a patient with rickettsialpox caused by *Rickettsia akari*. **B.** Papulovesicular lesions on the trunk of the same patient. **C.** Close-up of lesions from the same patient. (Reprinted from A Krusell et al: *Emerg Infect Dis* 8:727, 2002.)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-34

Ecthyma gangrenosum in a neutropenic patient with *Pseudomonas aeruginosa* bacteremia.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-35

Urticaria showing characteristic discrete and confluent, edematous, erythematous papules and plaques. (Reprinted from K Wolff, RA Johnson, AP Saavedra: Color Atlas and Synopsis of Clinical Dermatology, 7th ed. New York,



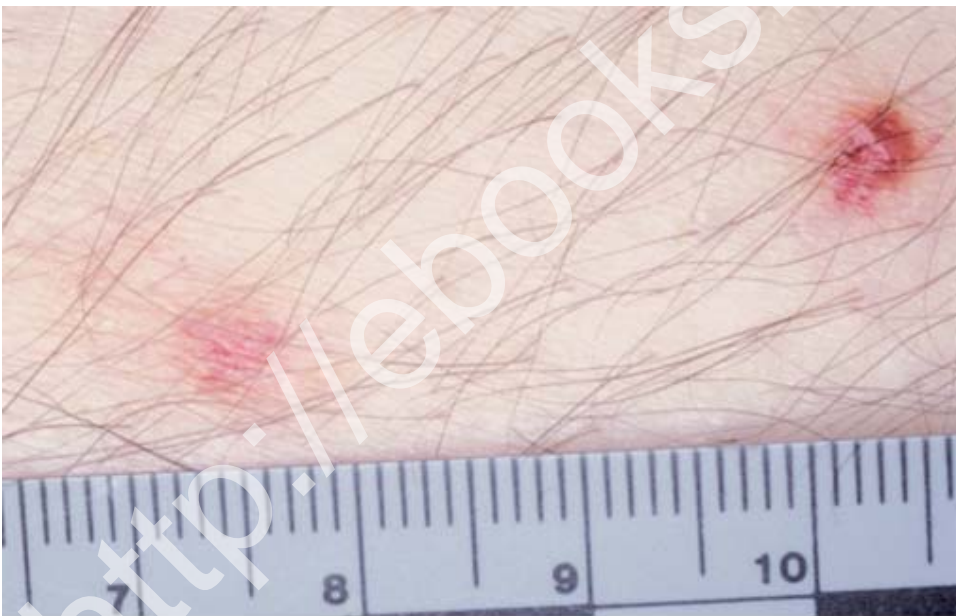
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-36

Disseminated cryptococcal infection. A liver transplant recipient developed six cutaneous lesions similar to the one shown. Biopsy and serum antigen testing demonstrated *Cryptococcus*. Important features of the lesion include a benign-appearing fleshy papule with central umbilication resembling molluscum contagiosum. (Courtesy of Lindsey Baden, MD; with permission.)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-37

Disseminated candidiasis. Tender, erythematous, nodular lesions developed in a neutropenic patient with leukemia who was undergoing induction chemotherapy. (*Courtesy of Lindsey Baden, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-38

Disseminated *Aspergillus* infection. Multiple necrotic lesions developed in this neutropenic patient undergoing hematopoietic stem cell transplantation. The lesion in the photograph is on the inner thigh and is several centimeters in diameter. Biopsy demonstrated infarction caused by *Aspergillus fumigatus*. (*Courtesy of Lindsey Baden, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-39

Erythema nodosum is a panniculitis characterized by tender, deep-seated nodules and plaques usually located on the lower extremities. (*Courtesy of Robert Swerlick, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-40

Sweet syndrome is an erythematous indurated plaque with a pseudovesicular border. (*Courtesy of Robert Swerlick, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-41

Fulminant meningococemia with extensive angular purpuric patches. (*Courtesy of Stephen E. Gellis, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-42

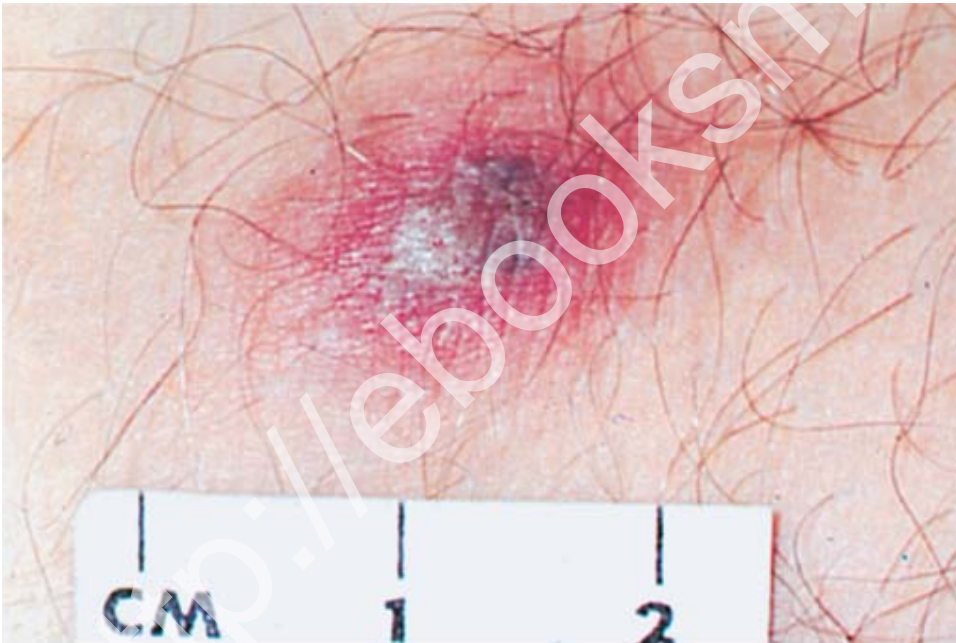
Erythematous papular lesions are seen on the leg of this patient with chronic meningococemia (*arrow indicates a lesion*).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-43

Disseminated gonococcemia in the skin is seen as hemorrhagic papules and pustules with purpuric centers, typically in a peripheral distribution near joints. (*Courtesy of Daniel M. Musher, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-44

Palpable purpuric papules on the lower leg are seen in this patient with cutaneous small-vessel hypersensitivity vasculitis. (*Reprinted from K Wolff, RA Johnson: Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-45

The thumb of a patient with a necrotic ulcer of tularemia. (*Courtesy of the Centers for Disease Control and Prevention.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-46

This 50-year-old man developed high fever and massive inguinal lymphadenopathy after a small ulcer healed on his foot. Tularemia was diagnosed. (*Courtesy of Lindsey Baden, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A1-47

This painful trypanosomal chancre developed at the site of a tsetse fly bite on the dorsum of the foot. *Trypanosoma brucei* was diagnosed from an aspirate of the ulcer. (*Courtesy of Edward T. Ryan, MD. N Engl J Med 346:2069, 2002; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-48

Drug reaction with eosinophilia and systemic symptoms/drug-induced hypersensitivity syndrome (DRESS/DIHS).

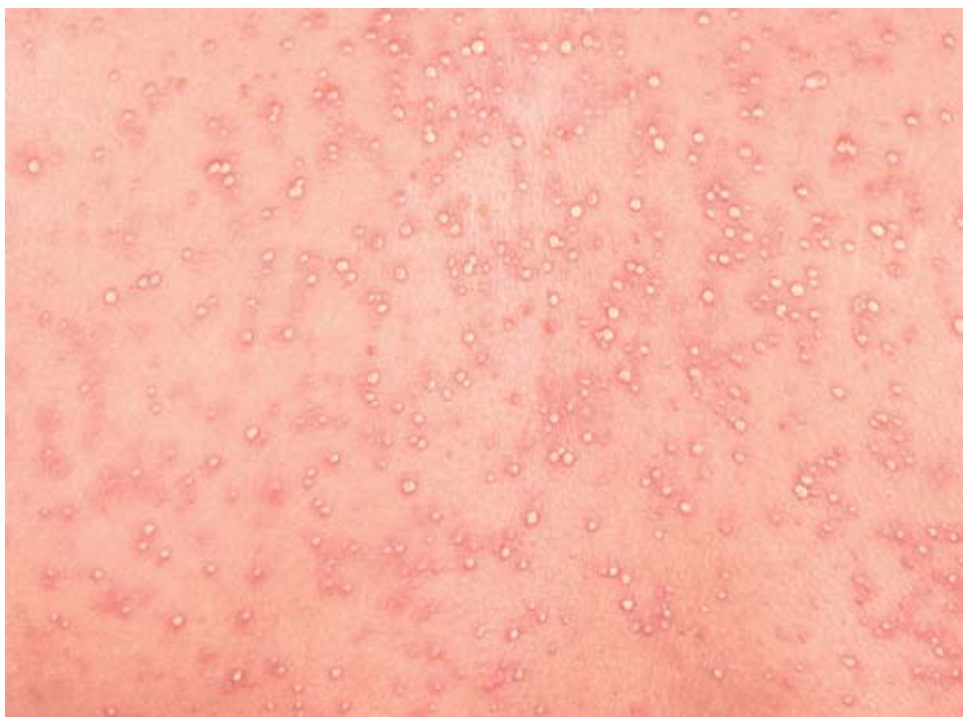
This patient developed a progressive eruption exhibiting early desquamation after taking phenobarbital. There was also associated lymphadenopathy and hepatomegaly. (*Courtesy of Peter Lio, MD; with permission.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-49

Many small, nonfollicular pustules are seen against a background of erythema with acute generalized exanthematous pustulosis (AGEP), typically resulting from a drug reaction. The rash began in body folds and progressed to cover the trunk and face. (*Reprinted from K Wolff, RA Johnson: Color Atlas and Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-50

Smallpox is shown with many pustules on the face, becoming confluent (**A**), and on the trunk (**B**). Pustules are all in the same stage of development. **C**. Crusting, healing lesions are noted on the trunk, arms, and hands. (Reprinted from K Wolff, RA Johnson: *Color Atlas and Synopsis of Clinical Dermatology*, 6th ed. New York, McGraw-Hill, 2009.)

**A****B****C**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: *Harrison's Principles of Internal Medicine*, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A1-51

Zika virus infection is shown with erythematous macules and papules on the arm and trunk (**A**) and on the foot (**B**). This patient also had conjunctival injection (**C**) and palatal petechiae (**D**). (Courtesy of Amit Garg, MD; with

permission.)



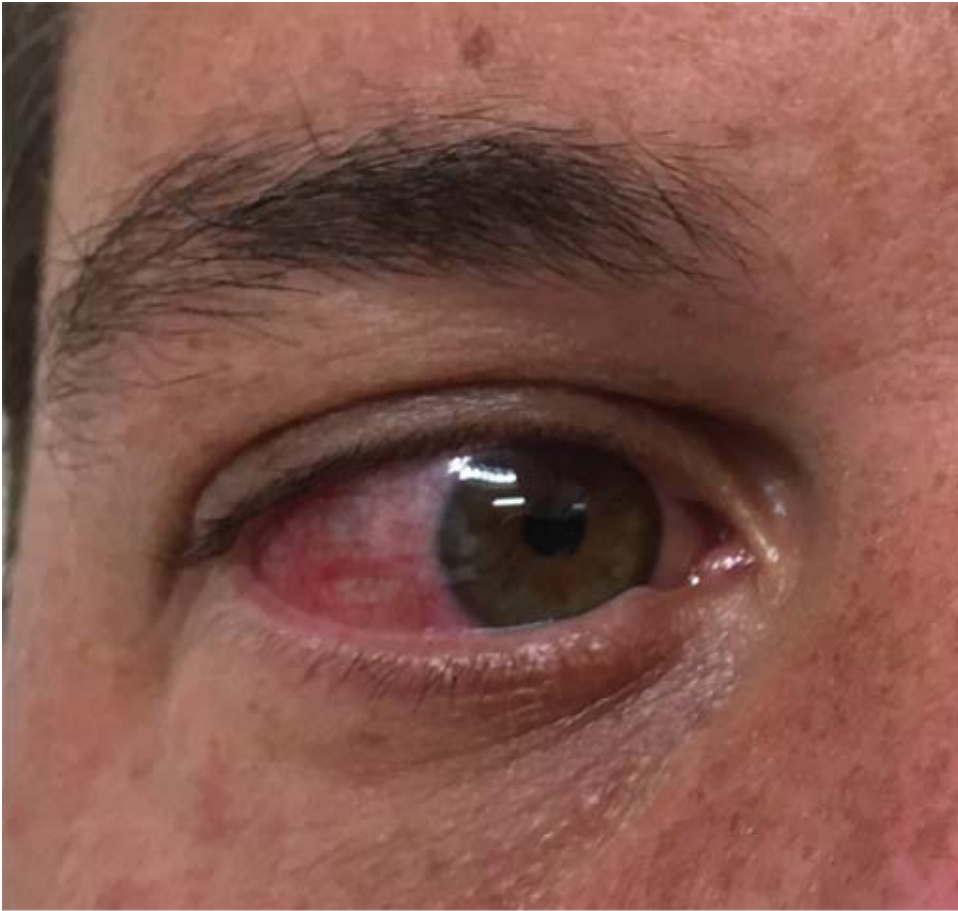
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



D

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

Chapter A2: Atlas of Oral Manifestations of Disease

Samuel C. Durso; Janet A. Yellowitz

The health status of the oral cavity is linked to cardiovascular disease, diabetes, and other systemic illnesses. Thus, examining the oral cavity for signs of disease is a key part of the physical examination. This chapter presents numerous outstanding clinical photographs illustrating many of the conditions discussed in [Chap. 32](#), Oral Manifestations of Disease. Conditions affecting the teeth, periodontal tissues, and oral mucosa are all represented.

FIGURE A2-1

Gingival overgrowth secondary to calcium channel blocker use.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A2-2

Oral lichen planus.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A2-3

Erosive lichen planus.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A2-4

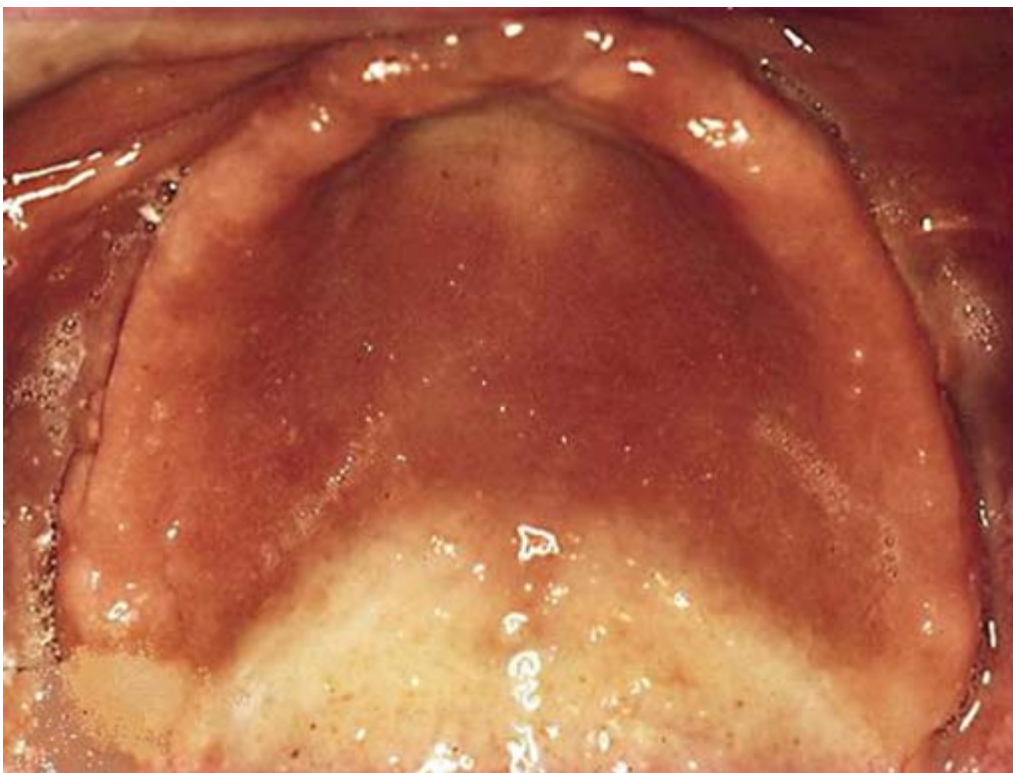
Stevens-Johnson syndrome—reaction to nevirapine.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A2-5

Erythematous candidiasis under a denture (i.e., the patient should be treated for this fungal infection).



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A2-6

Severe periodontitis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-7

Angular cheilitis.

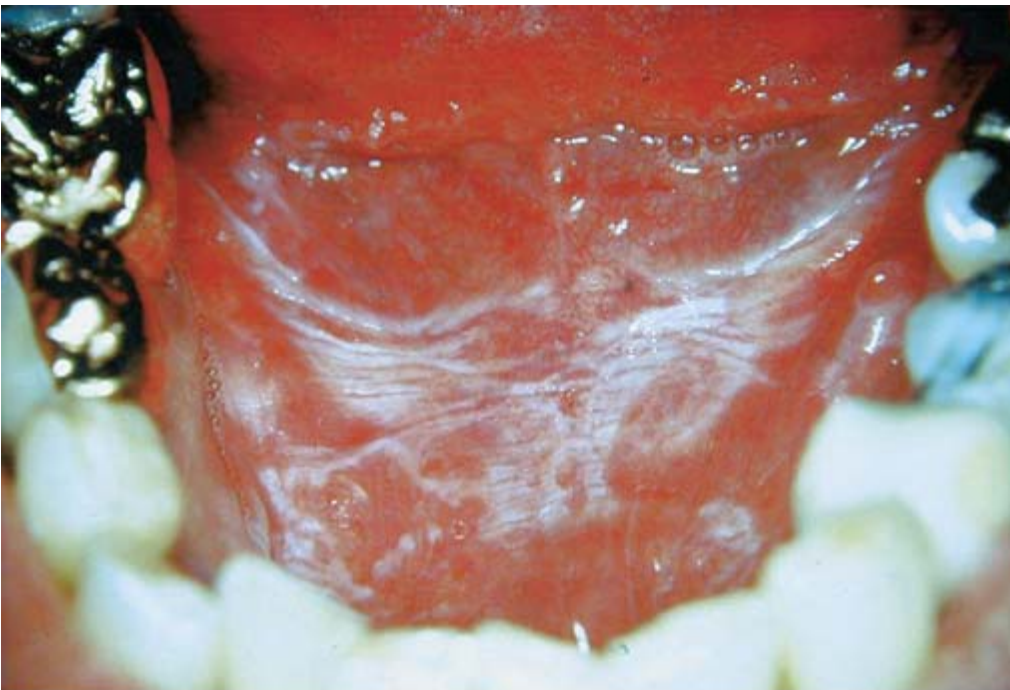


Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-8

Sublingual leukoplakia.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-9

A. Epulis (gingival hypertrophy) under denture. B. Epulis fissuratum.



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

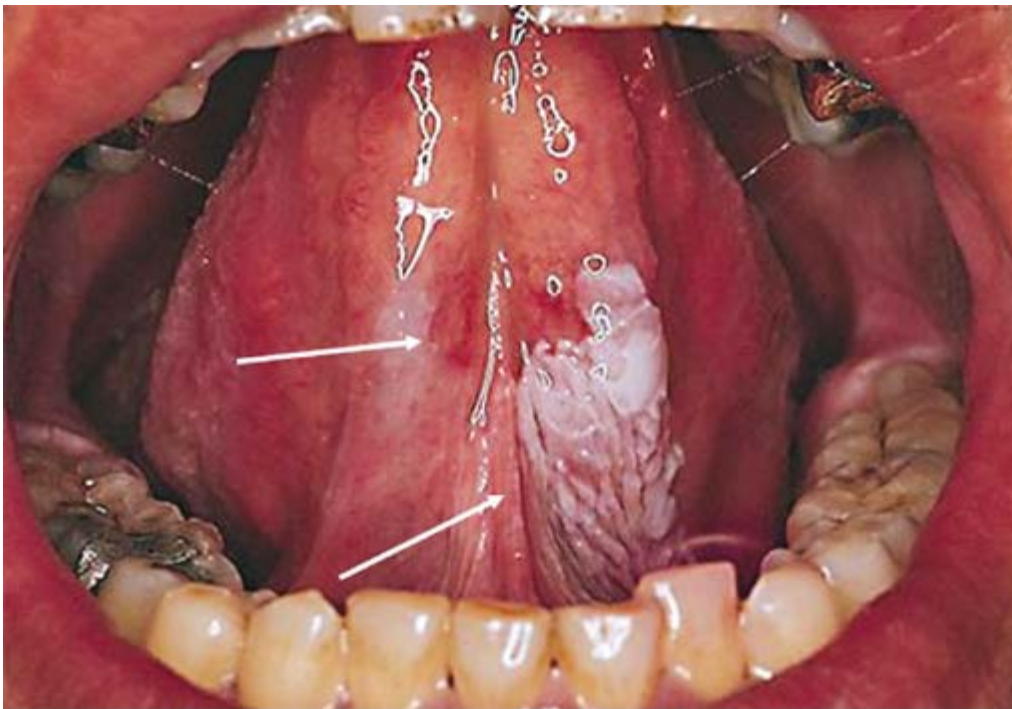
FIGURE A2-10

Traumatic lesion inside of cheek.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-11

Oral leukoplakia, subtype homogenous leukoplakia.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-12

Oral carcinoma.



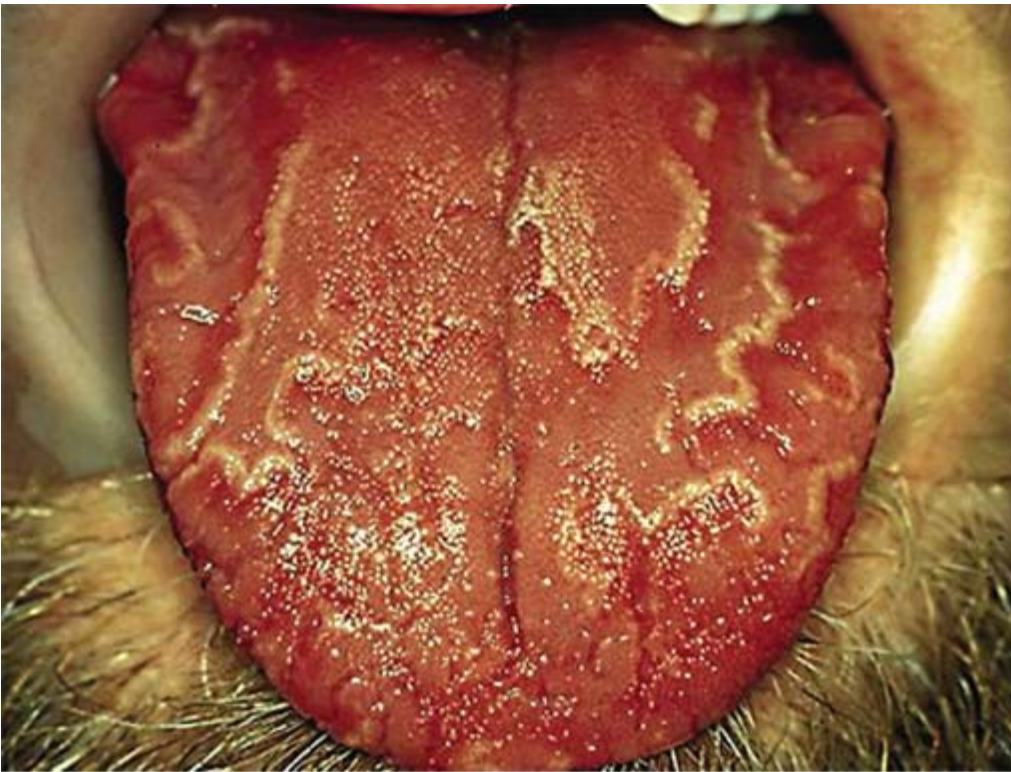
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-13

Healthy mouth.



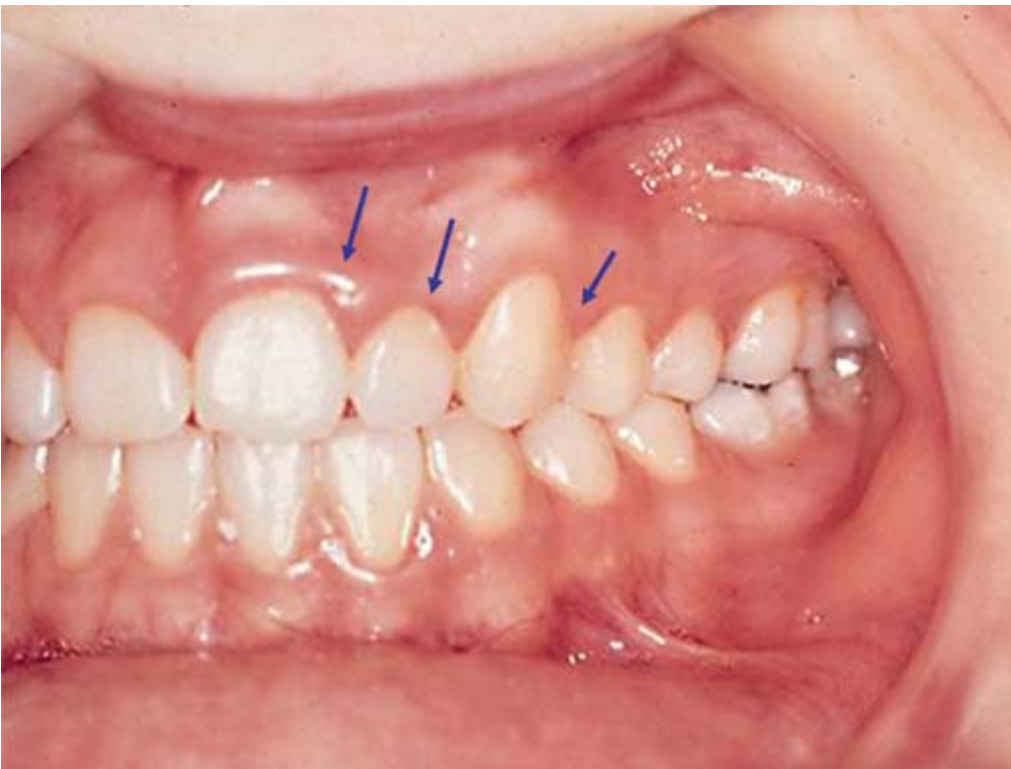
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-14

Geographic tongue.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-15

Moderate gingivitis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-16

Gingival recession.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-17

Heavy calculus and gingival inflammation.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-18

Severe gingival inflammation and heavy calculus.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-19

Root cavity in presence of severe periodontal disease.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-20

Ulcer on lateral border of tongue—potential carcinoma.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-21

Osteonecrosis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-22

Severe periodontal disease, missing tooth, very mobile teeth.



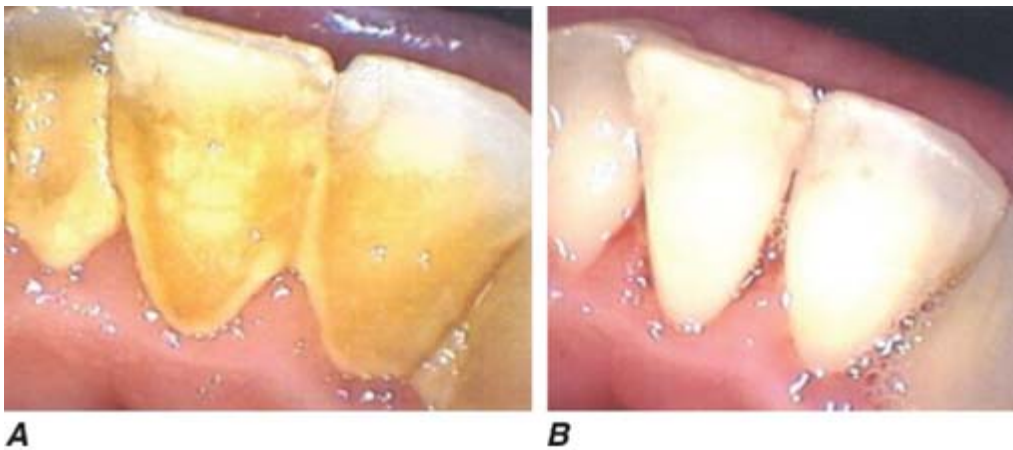
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-23

Salivary stone.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-24

A. Calculus. B. Teeth cleaned.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-25

Traumatic ulcer.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-26

Fissured tongue.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A2-27

White coated tongue—likely candidiasis.



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

Acknowledgment

Dr. Jane Atkinson was a co-author of this chapter in the 17th edition. Some of the materials have been carried over into this edition.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)

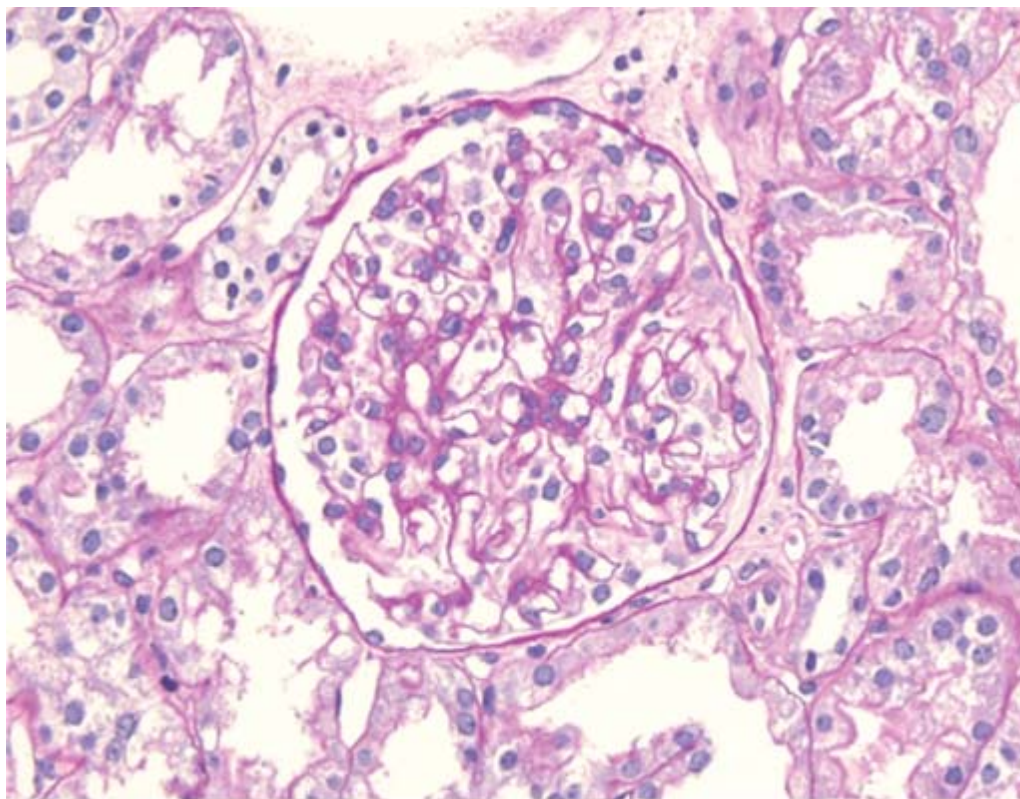
Chapter A3: Atlas of Urinary Sediments and Renal Biopsies

Agnes B. Fogo; Eric G. Neilson

Key diagnostic features of selected diseases in renal biopsy are illustrated, with light, immunofluorescence, and electron microscopic images. Common urinalysis findings are also documented.

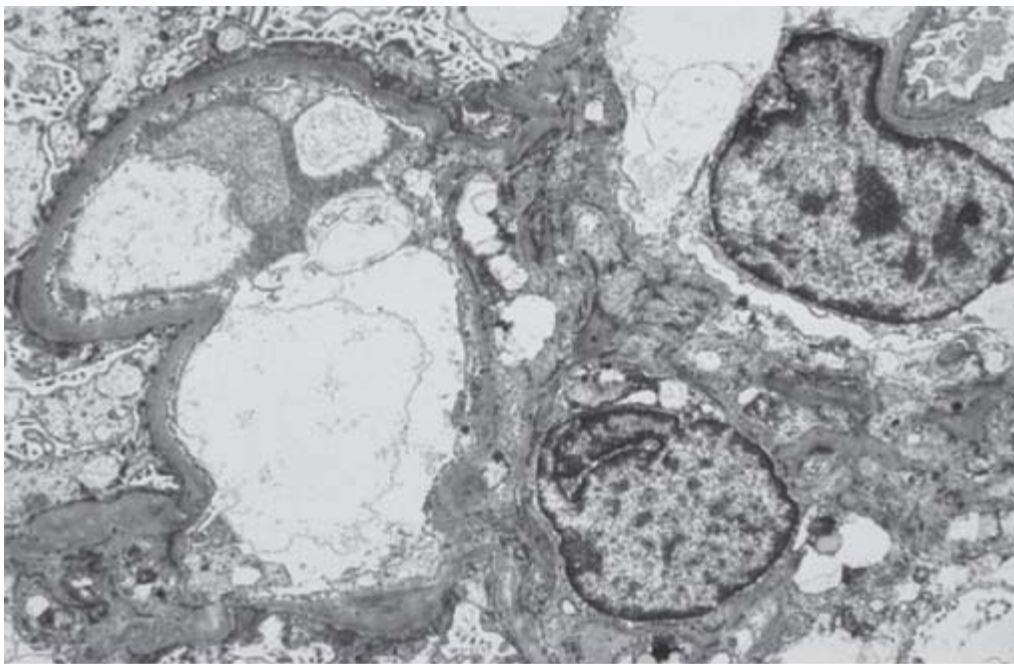
FIGURE A3-1

Minimal change disease. In minimal change disease, light microscopy is unremarkable (**A**), whereas electron microscopy (**B**) reveals podocyte injury evidenced by complete foot process effacement. (*ABF/Vanderbilt Collection.*)



A

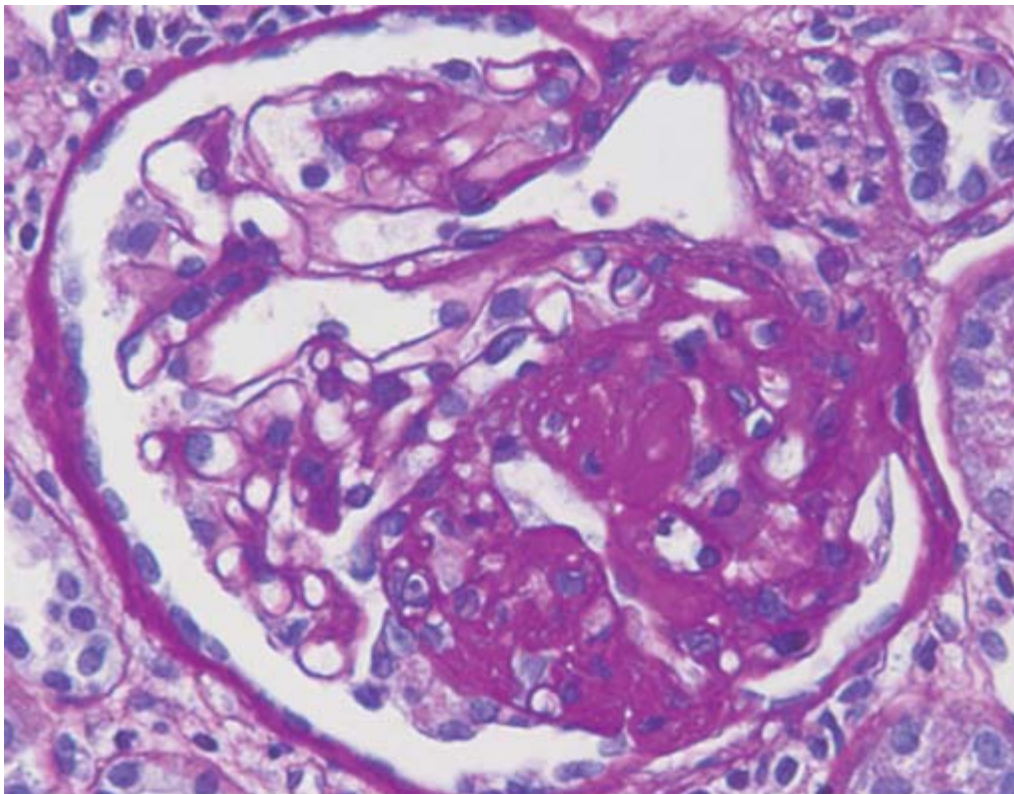
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-2

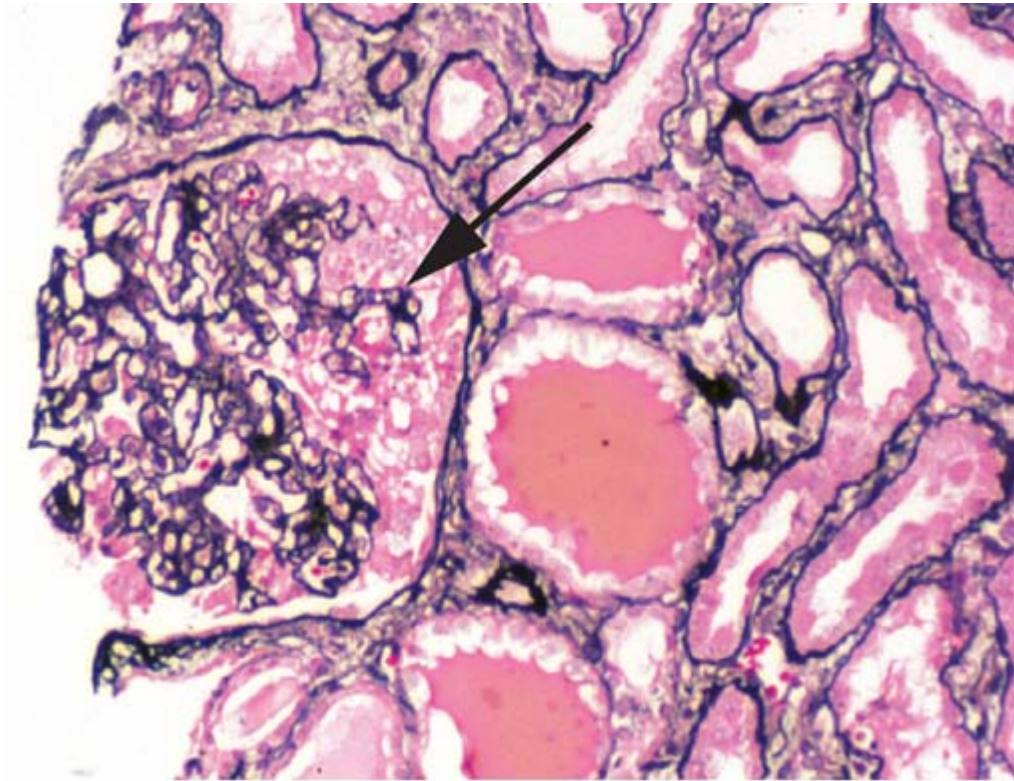
Focal segmental glomerulosclerosis (FSGS). There is a well-defined segmental increase in matrix and obliteration of capillary loops in the right part of the glomerulus, the sine qua non of segmental sclerosis not otherwise specified (NOS) type. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-3

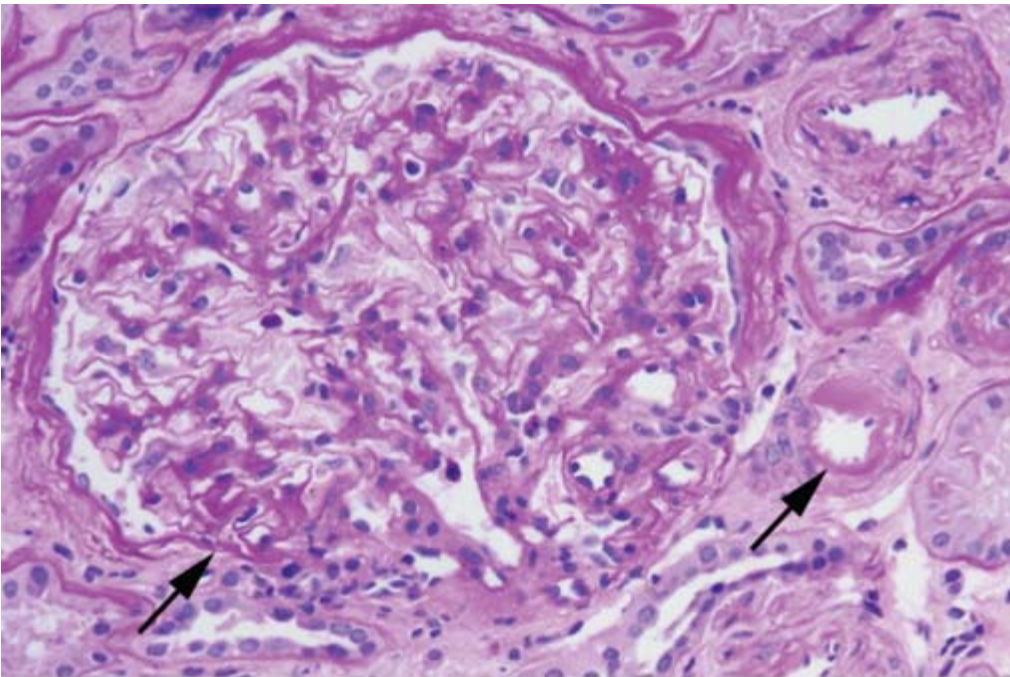
Collapsing glomerulopathy. There is segmental collapse (*arrow*) of the glomerular capillary loops and overlying podocyte hyperplasia. This lesion may be idiopathic or associated with e.g., HIV infection and has a particularly poor prognosis. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

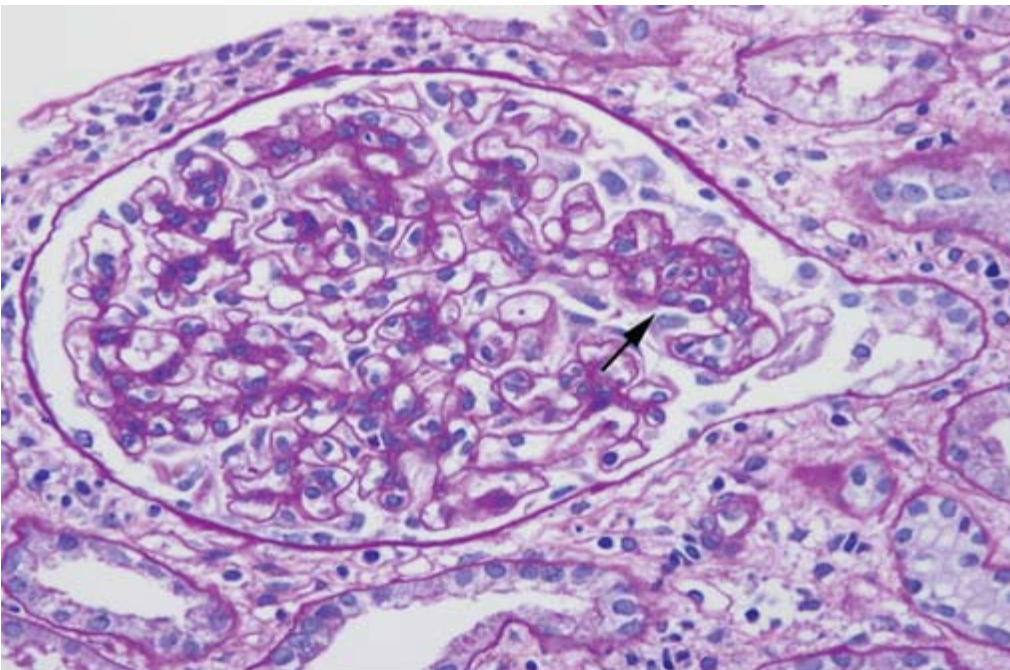
FIGURE A3-4

Hilar variant of FSGS. There is segmental sclerosis of the glomerular tuft at the vascular pole with associated hyalinosis, also present in the afferent arteriole (*arrows*). This lesion often occurs as a secondary response when nephron mass is lost due to, e.g., scarring from other conditions. Patients usually have less proteinuria and less steroid response than FSGS, NOS type. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-5

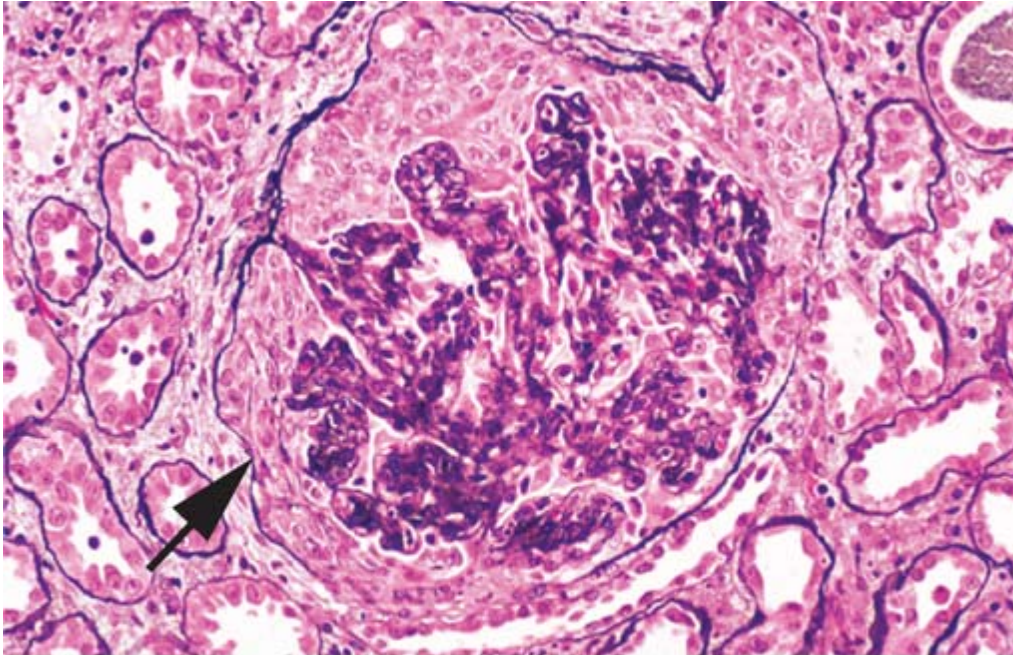
Tip lesion variant of FSGS. There is segmental sclerosis of the glomerular capillary loops at the proximal tubular outlet (*arrow*). This lesion has a better prognosis than other types of FSGS. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-6

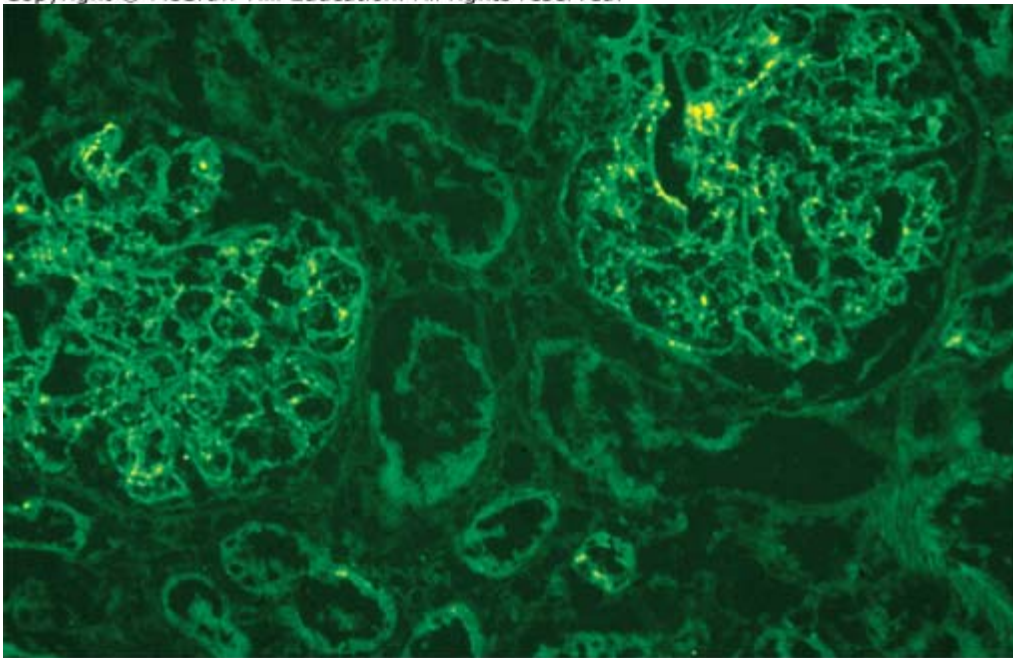
Postinfectious (poststreptococcal) glomerulonephritis. The glomerular tuft shows proliferative changes with numerous polymorphonuclear leukocytes (PMNs), with a crescentic reaction (*arrow*) in severe cases (**A**). These deposits localize in the mesangium and along the capillary wall in a subepithelial pattern and stain

dominantly for C3 and to a lesser extent for IgG (**B**). Subepithelial hump-shaped deposits are seen by electron microscopy (*arrow*) (**C**). (*ABF/Vanderbilt Collection.*)



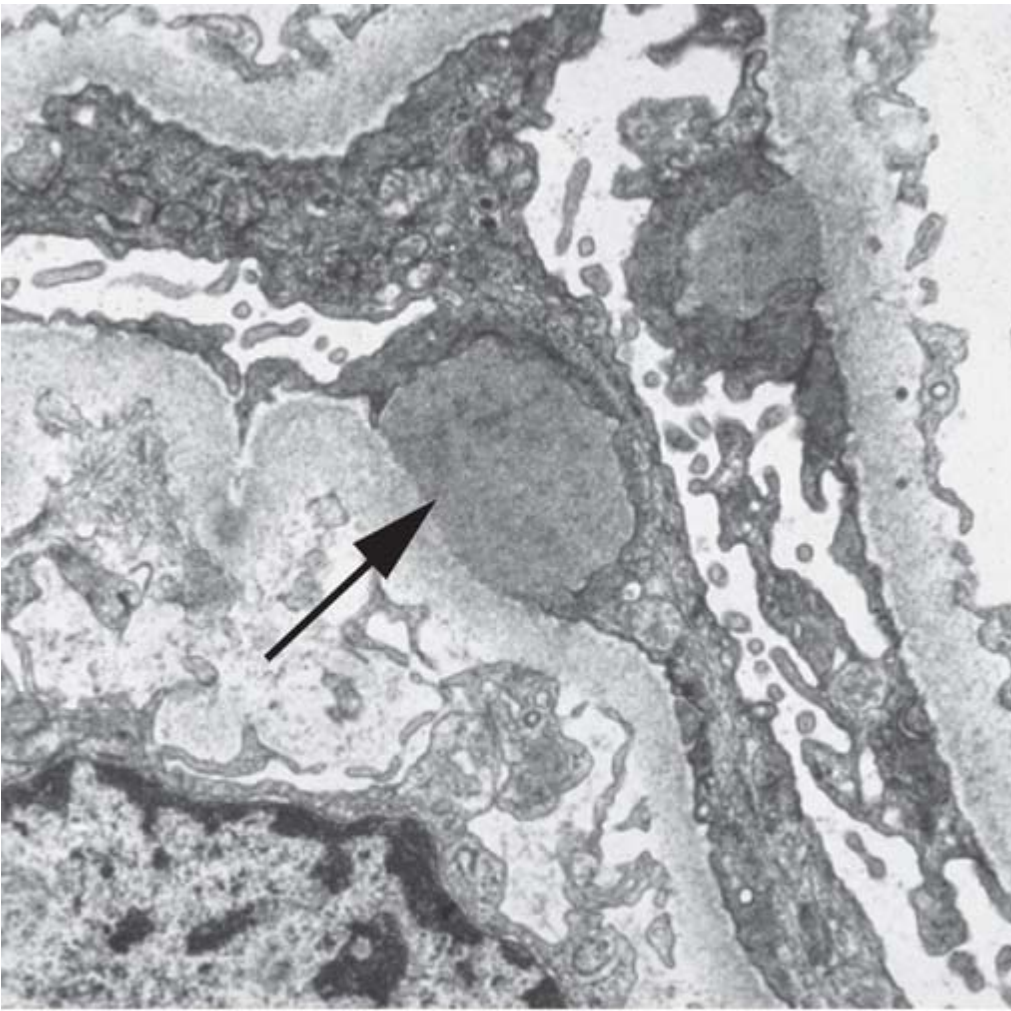
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

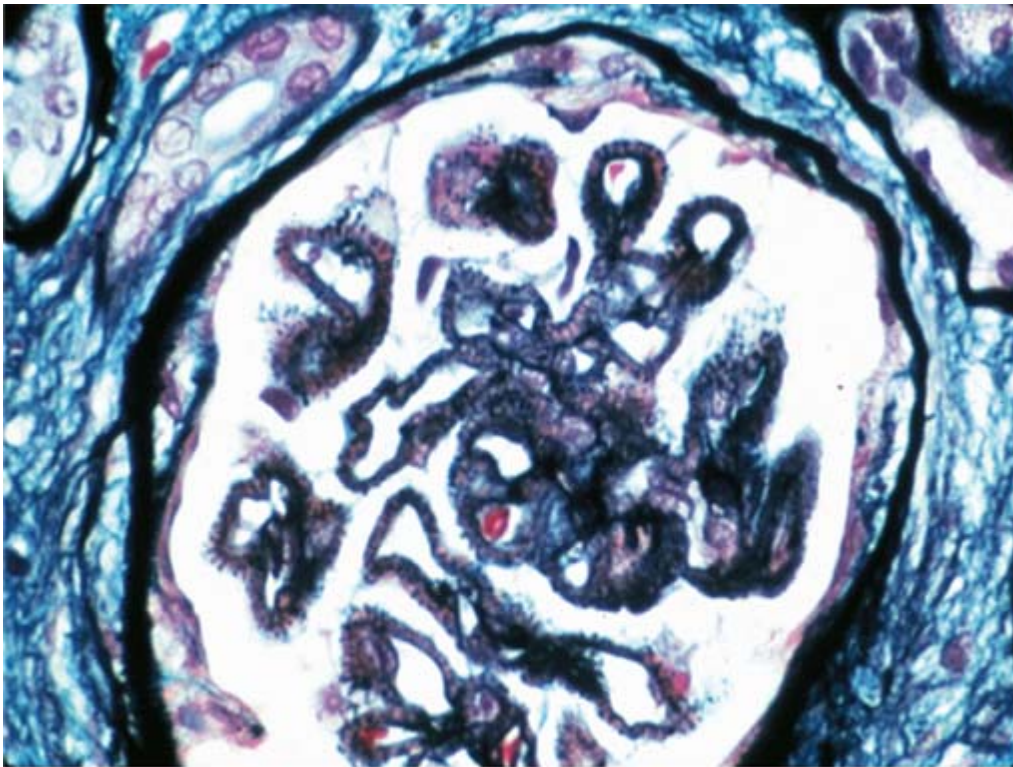


C

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

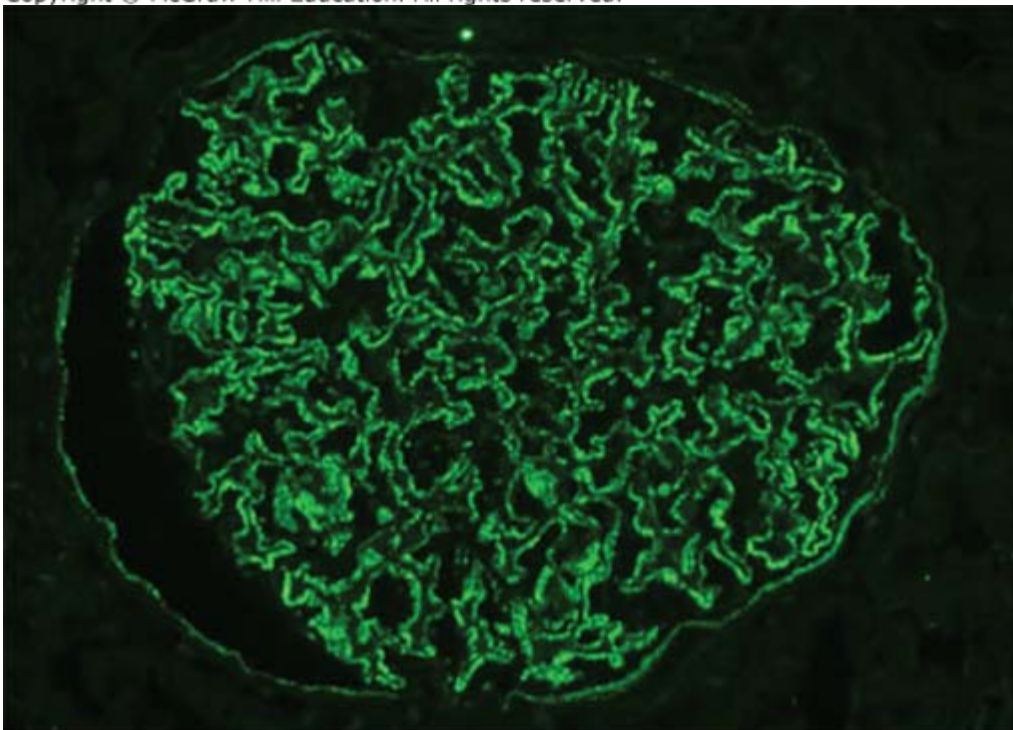
FIGURE A3-7

Membranous nephropathy. Membranous nephropathy is due to subepithelial deposits, with resulting basement membrane reaction, resulting in the appearance of spike-like projections on silver stain (**A**). The deposits are directly visualized by fluorescent anti-IgG, revealing diffuse granular capillary loop staining (**B**). By electron microscopy, the subepithelial location of the deposits and early surrounding basement membrane reaction are evident, with overlying foot process effacement (**C**). Most cases of primary membranous nephropathy are due to autoantibodies to the phospholipase A2 receptor (PLA2R), which is present on podocytes, and this antigen can then be detected in the deposits by staining with anti-PLA2R antibody (**D**). (ABF/Vanderbilt Collection.)



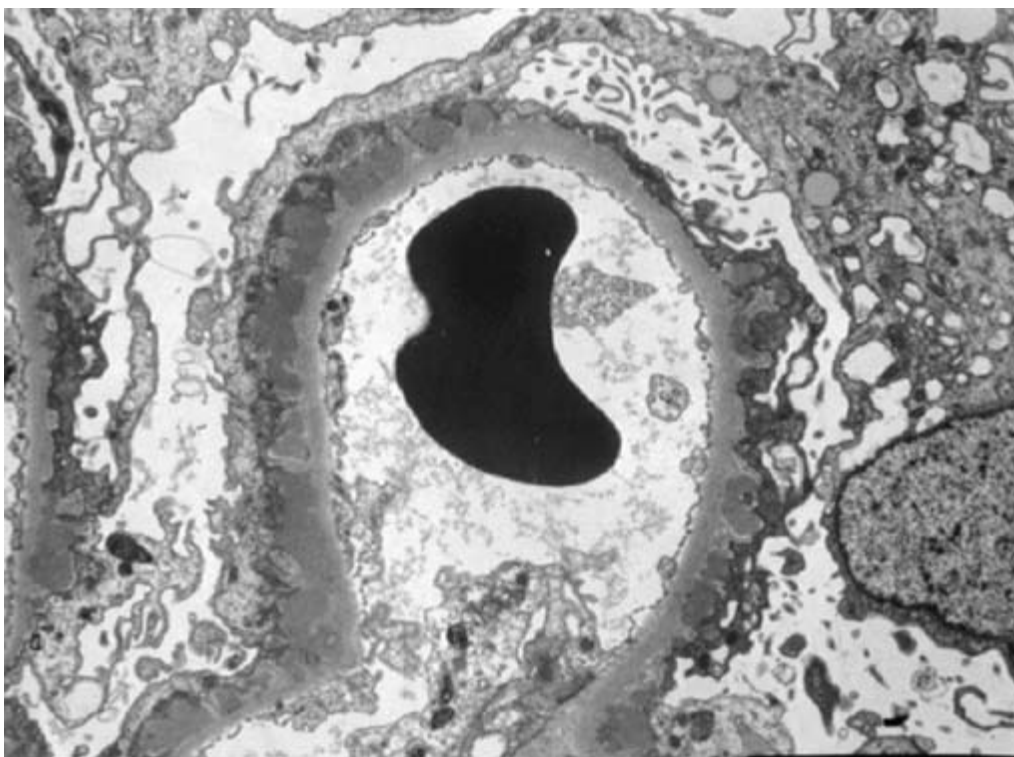
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



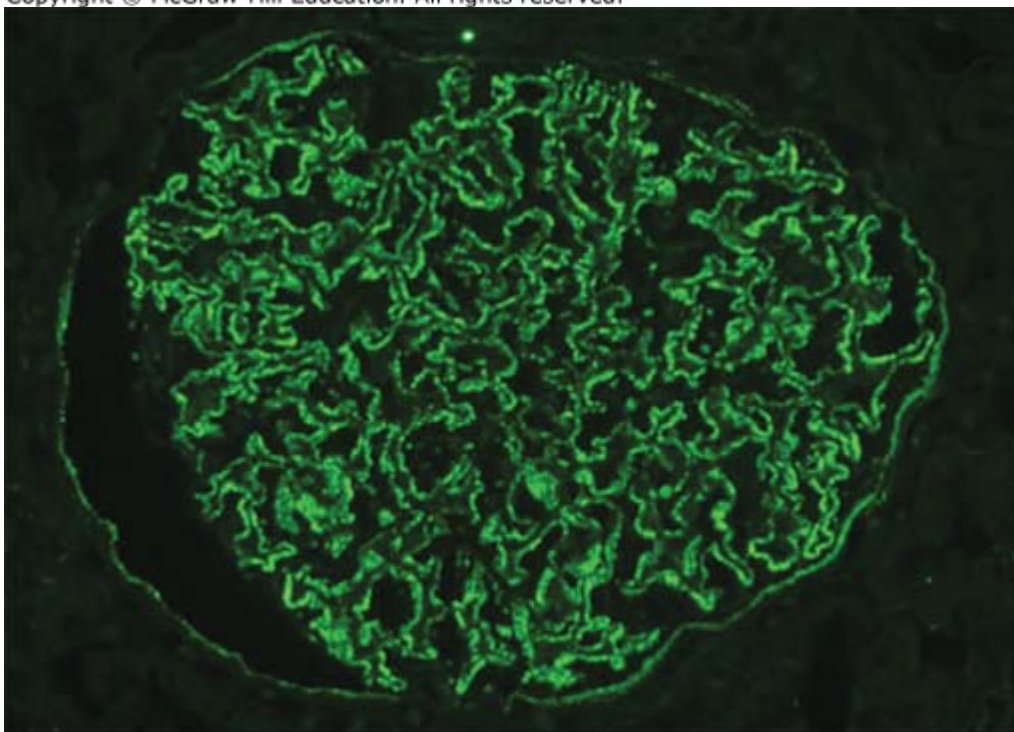
B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



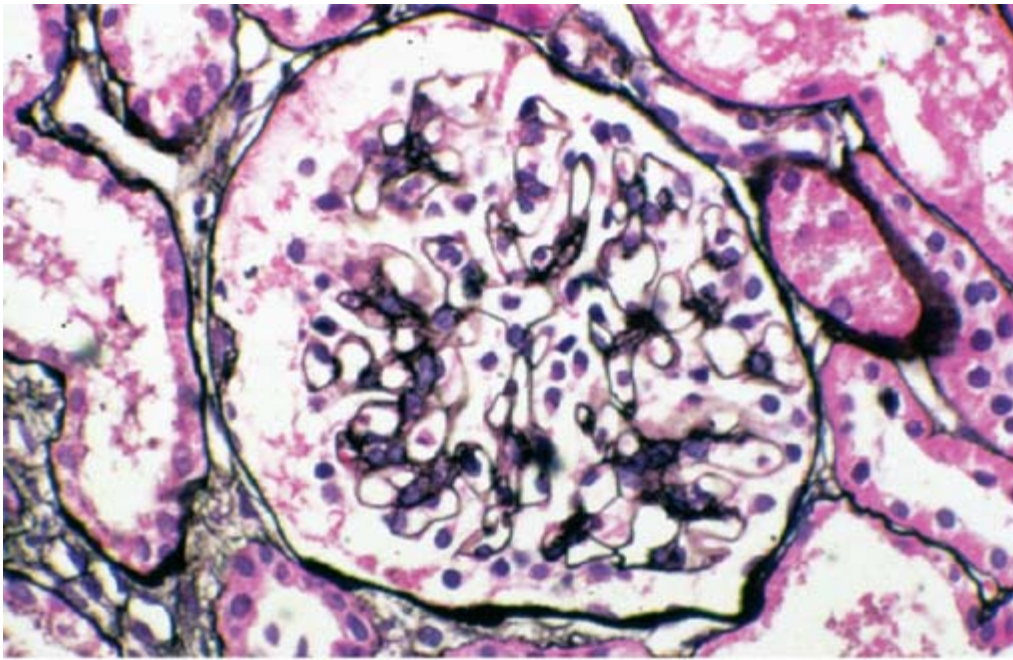
D

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-8

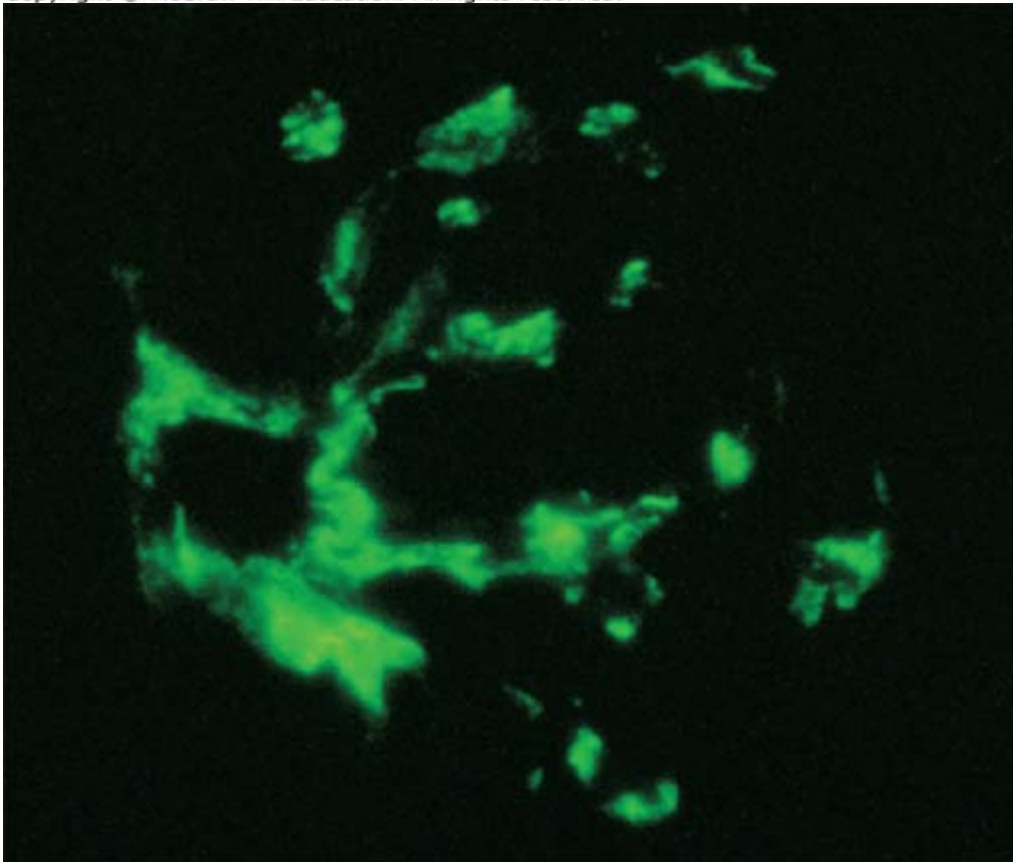
IgA nephropathy. There is variable mesangial expansion due to mesangial deposits, with some cases also showing endocapillary hypercellularity or segmental sclerosis (**A**). By immunofluorescence, mesangial IgA

deposits are evident (**B**). (ABF/Vanderbilt Collection.)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

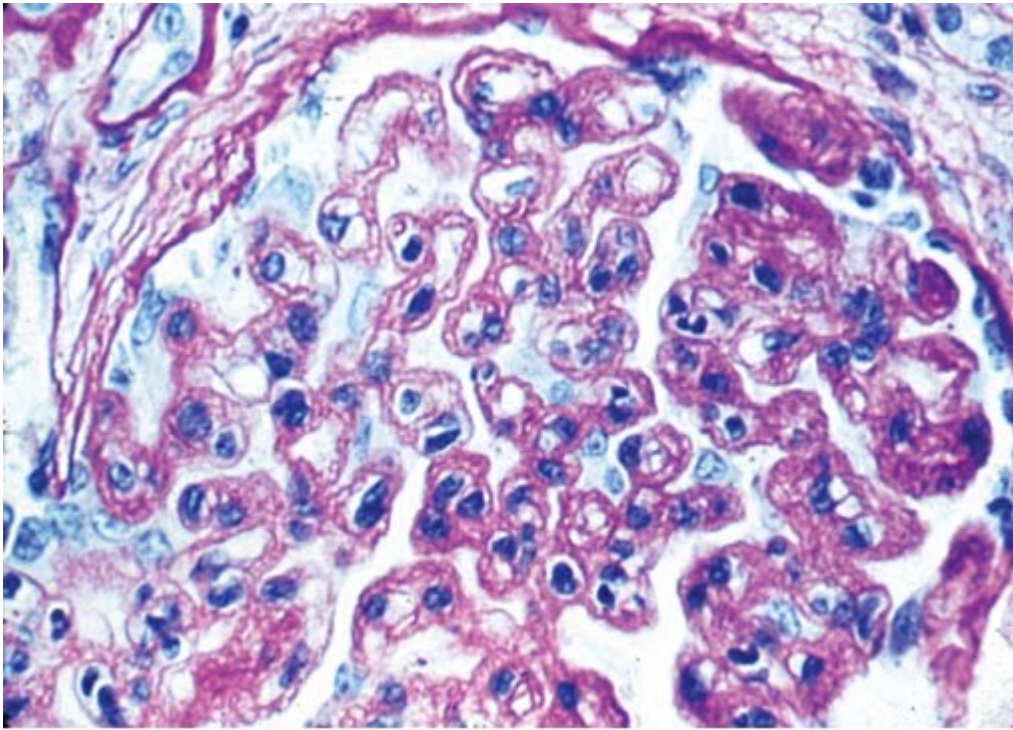


B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

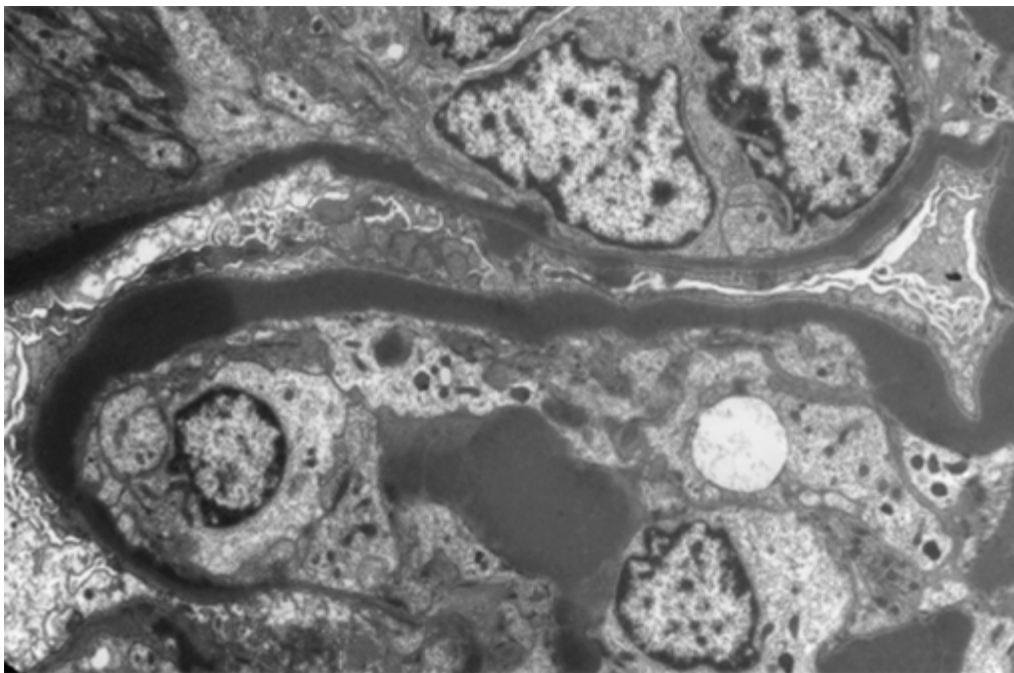
FIGURE A3-9

Membranoproliferative glomerulonephritis. There is mesangial expansion and endocapillary hypercellularity with cellular interposition in response to subendothelial deposits, resulting in the “tram-track” of duplication of glomerular basement membrane. (*EGN/UPenn Collection.*)



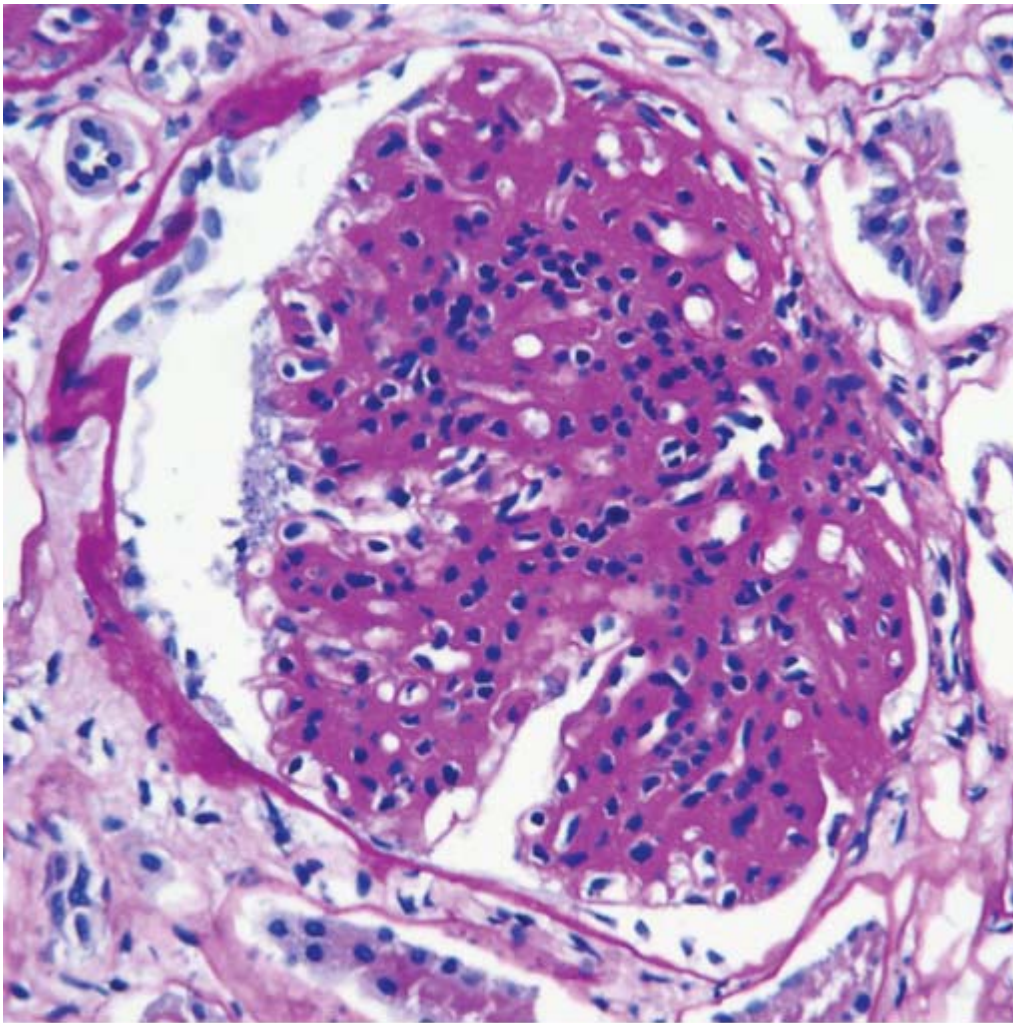
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-10

Dense deposit disease (membranoproliferative glomerulonephritis type II). By light microscopy, there is a membranoproliferative pattern. By electron microscopy, there is a dense transformation of the glomerular basement membrane with round, globular deposits within the mesangium. By immunofluorescence, only C3 staining is usually present. Dense deposit disease is part of the group of renal diseases called C3 glomerulopathy, related to underlying complement dysregulation. (*ABF/Vanderbilt Collection.*)



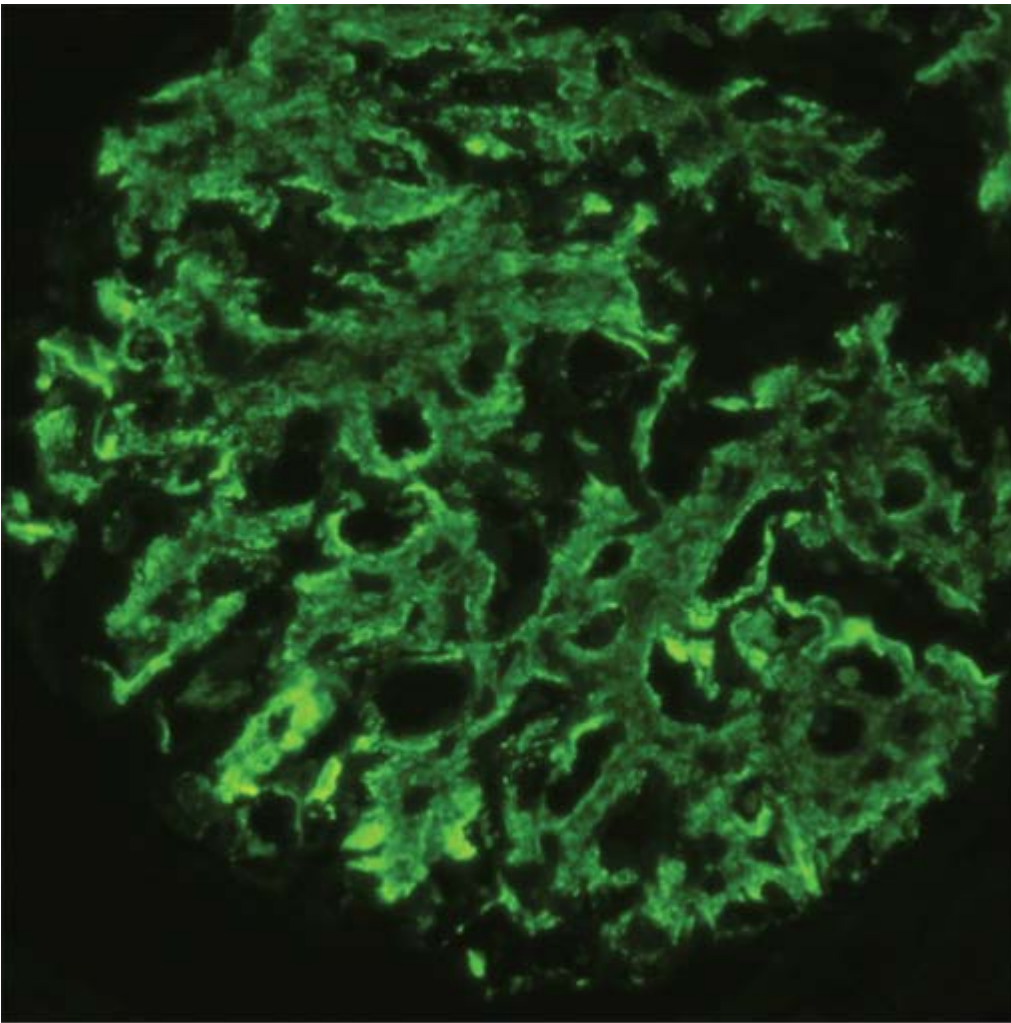
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-11

C3 glomerulonephritis. By light microscopy, there is a membranoproliferative pattern. C3 glomerulonephritis is part of the group of renal diseases called C3 glomerulopathy, related to underlying complement dysregulation. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-12

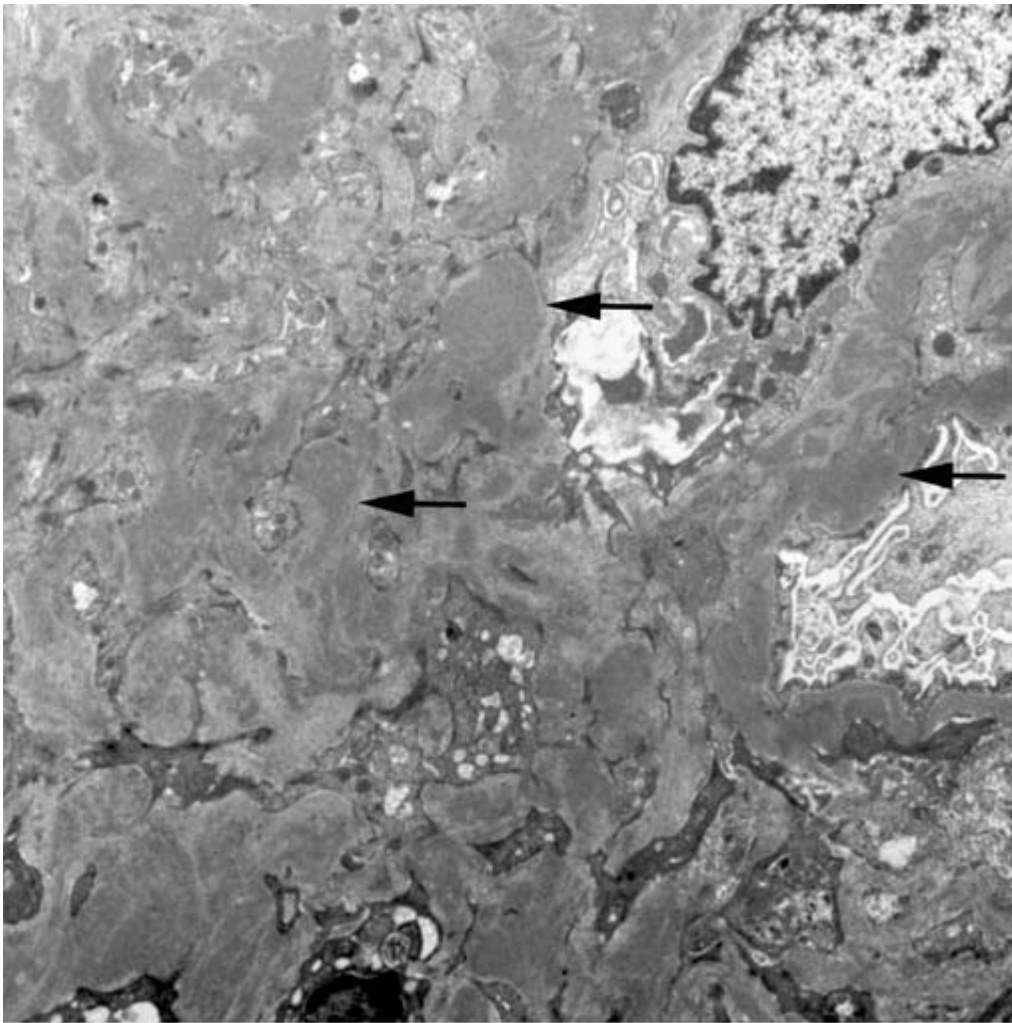
C3 glomerulonephritis. By immunofluorescence, only C3 staining is usually present, with occasional minimal immunoglobulin, in an irregular capillary wall and mesangial distribution. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-13

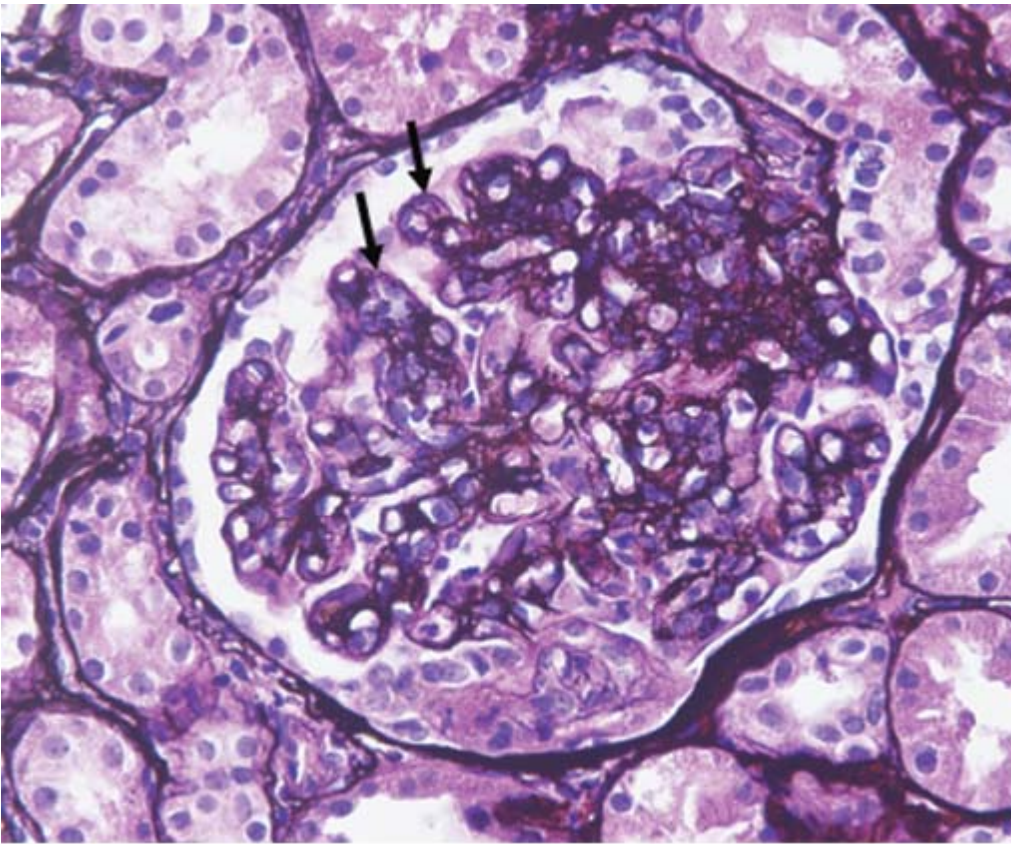
C3 glomerulonephritis. By electron microscopy, usual density deposits are present (*arrows*), including mesangial, subendothelial, and occasional hump-type subepithelial deposits. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

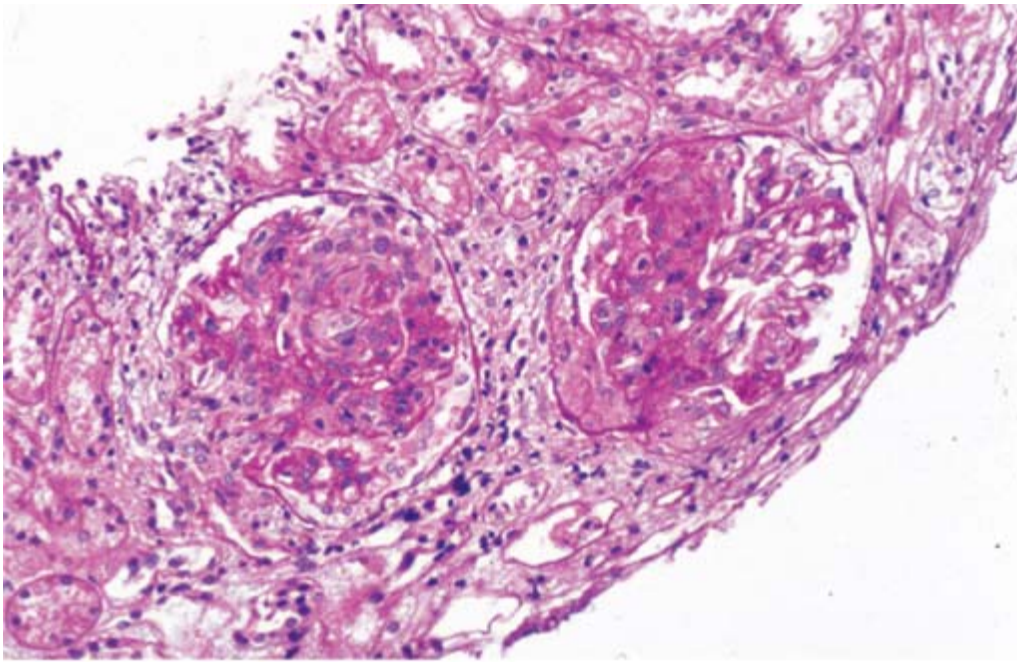
FIGURE A3-14

Lupus nephritis with mixed proliferative and membranous glomerulonephritis lesions. This specimen shows pink subepithelial deposits with spike reaction, and the “tram-track” sign of duplication of glomerular basement membrane, resulting from subendothelial deposits, as may be seen in mixed membranous and focal or diffuse lupus nephritis (International Society of Nephrology [ISN]/Renal Pathology Society [RPS] class V combined with class III or IV) Segmental endocapillary and mesangial hypercellularity are evident, along with spikes (lower arrow) and segmental double contours of the glomerular basement membrane (top arrow). A small cellular crescent is present at 6 o'clock. (*EGN/UPenn Collection/ABF/Vanderbilt Collection.*)



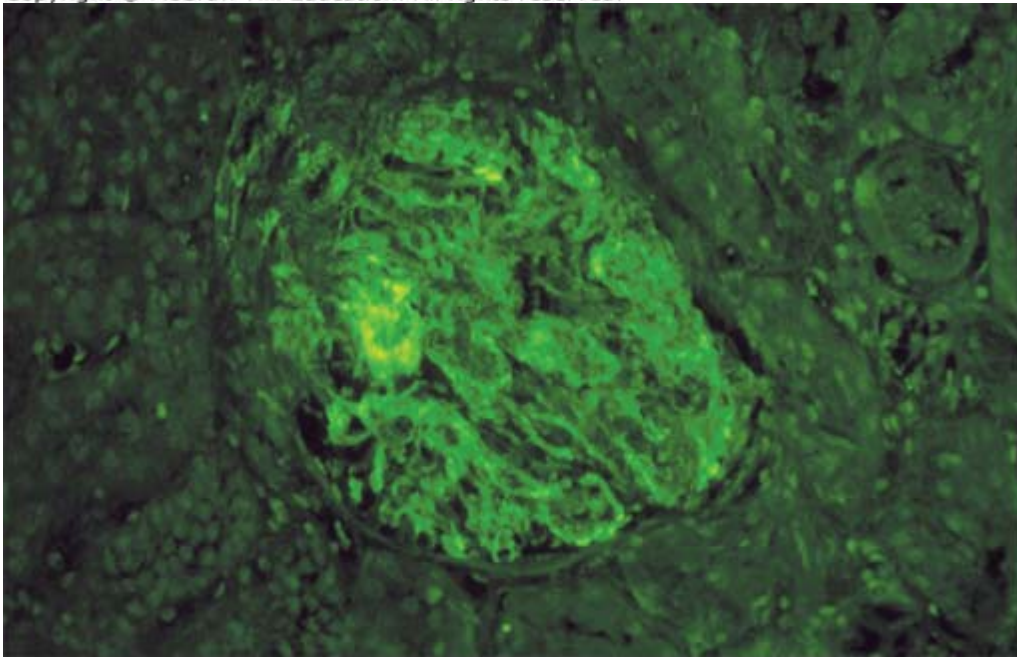
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-15

Lupus nephritis. Proliferative lupus nephritis, ISN/RPS class III (focal) or IV (diffuse), manifests as endocapillary hypercellularity, which may result in segmental necrosis due to deposits, particularly in the subendothelial area (**A**). By immunofluorescence, chunky irregular mesangial and capillary loop deposits are evident, with some of the peripheral loop deposits having a smooth, molded outer contour due to their subendothelial location. These deposits typically stain for all three immunoglobulins, IgG, IgA, IgM, and both C3 and C1q (**B**). By electron microscopy, subendothelial (*arrow*), mesangial (*white rim arrowhead*), and rare subepithelial (*black arrowhead*) immune complex deposits are evident, along with extensive foot process effacement (**C**). (*ABF/Vanderbilt Collection.*)



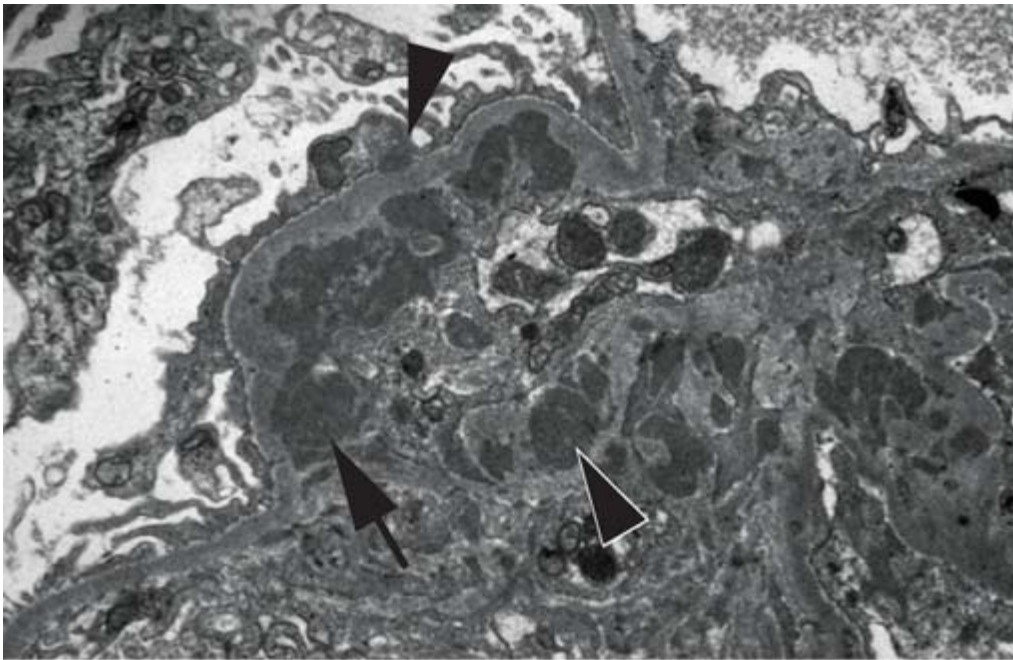
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

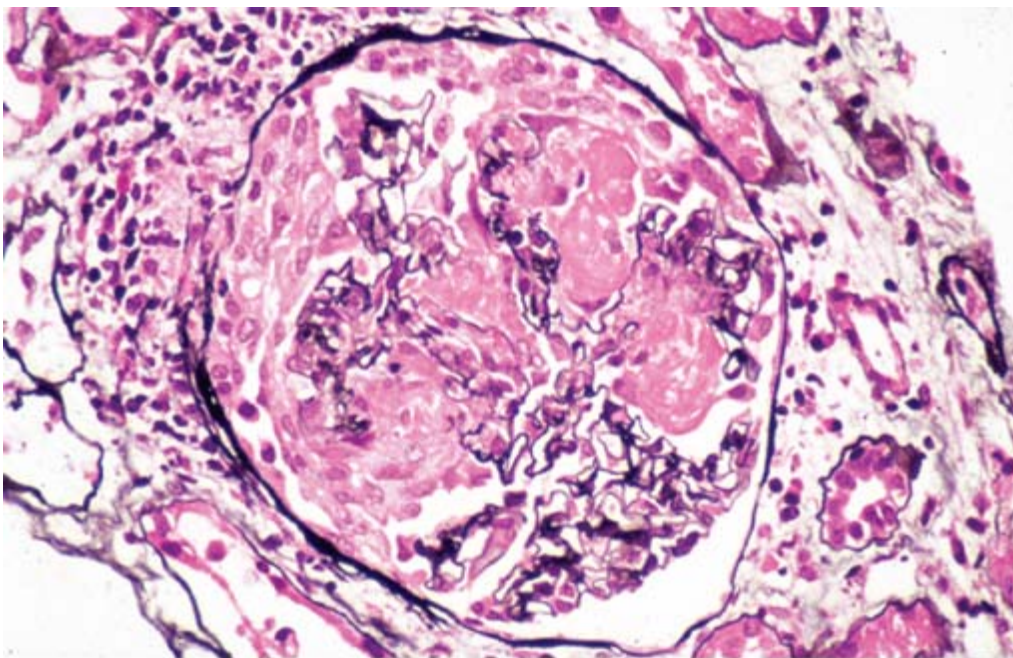
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

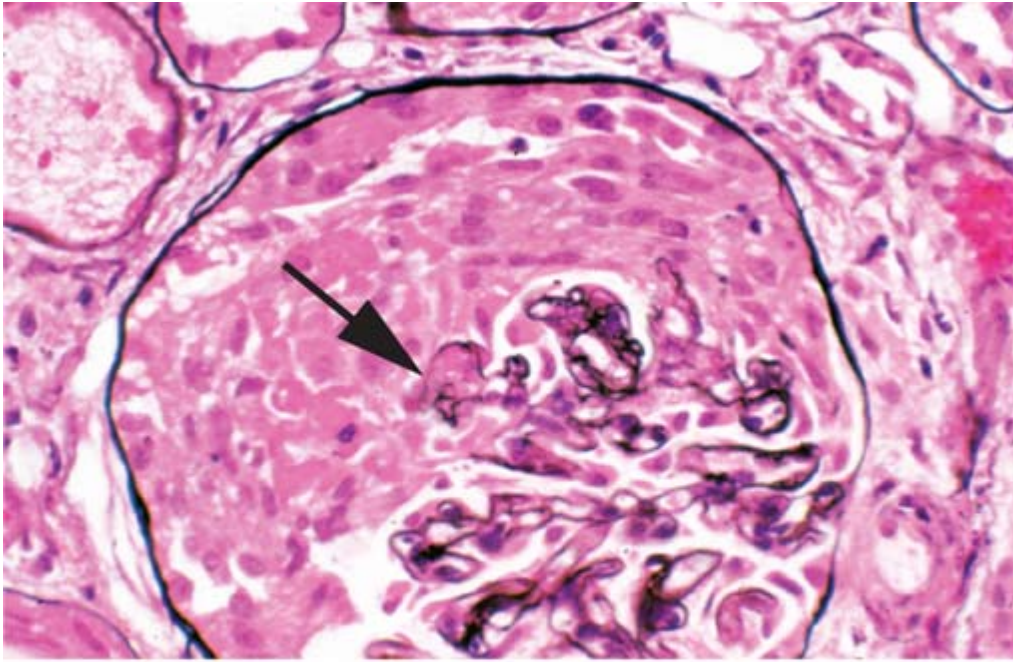
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
 Copyright © McGraw-Hill Education. All rights reserved.
 FIGURE A3-16

Granulomatosis with polyangiitis (Wegener's granulomatosis). This pauci-immune necrotizing crescentic glomerulonephritis shows numerous breaks in the glomerular basement membrane with associated segmental fibrinoid necrosis and a crescent formed by proliferation of the parietal epithelium. Note that the uninvolved segment of the glomerulus (at ~5 o'clock) shows no evidence of proliferation or immune complexes. (*ABF/Vanderbilt Collection.*)



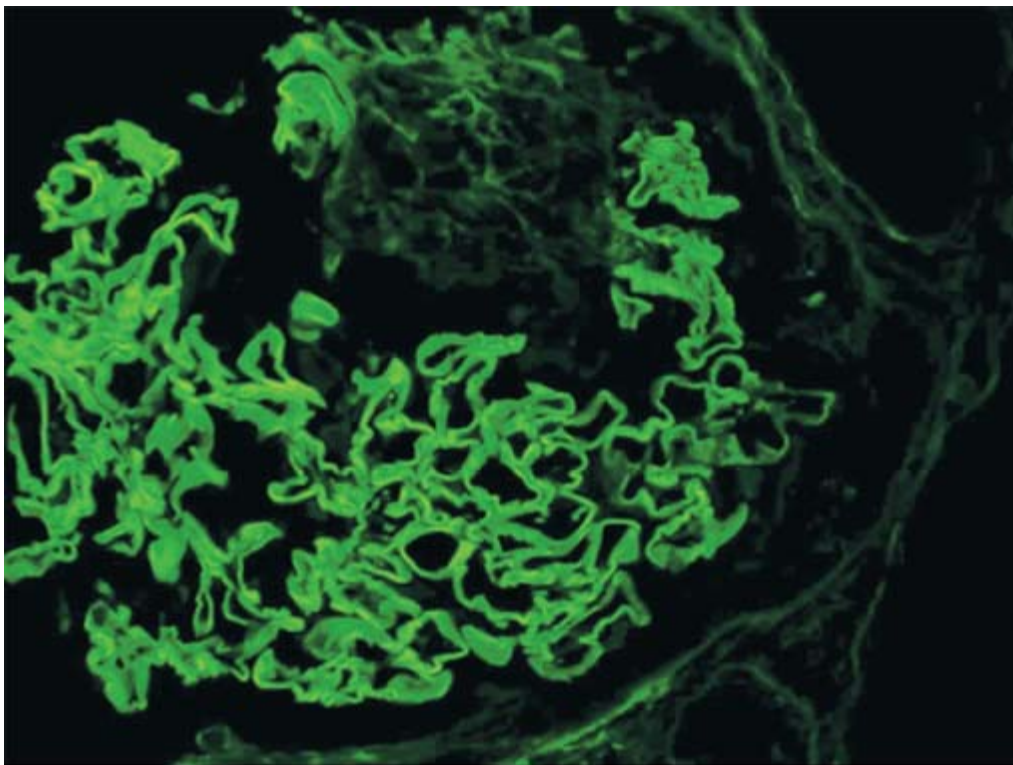
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
 Copyright © McGraw-Hill Education. All rights reserved.
 FIGURE A3-17

Anti-glomerular basement membrane antibody-mediated glomerulonephritis. There is segmental necrosis with a break of the glomerular basement membrane (*arrow*) and a cellular crescent (**A**), and immunofluorescence for IgG shows linear staining of the glomerular basement membrane with a small crescent at ~1 o'clock (**B**). (*ABF/Vanderbilt Collection.*)



A

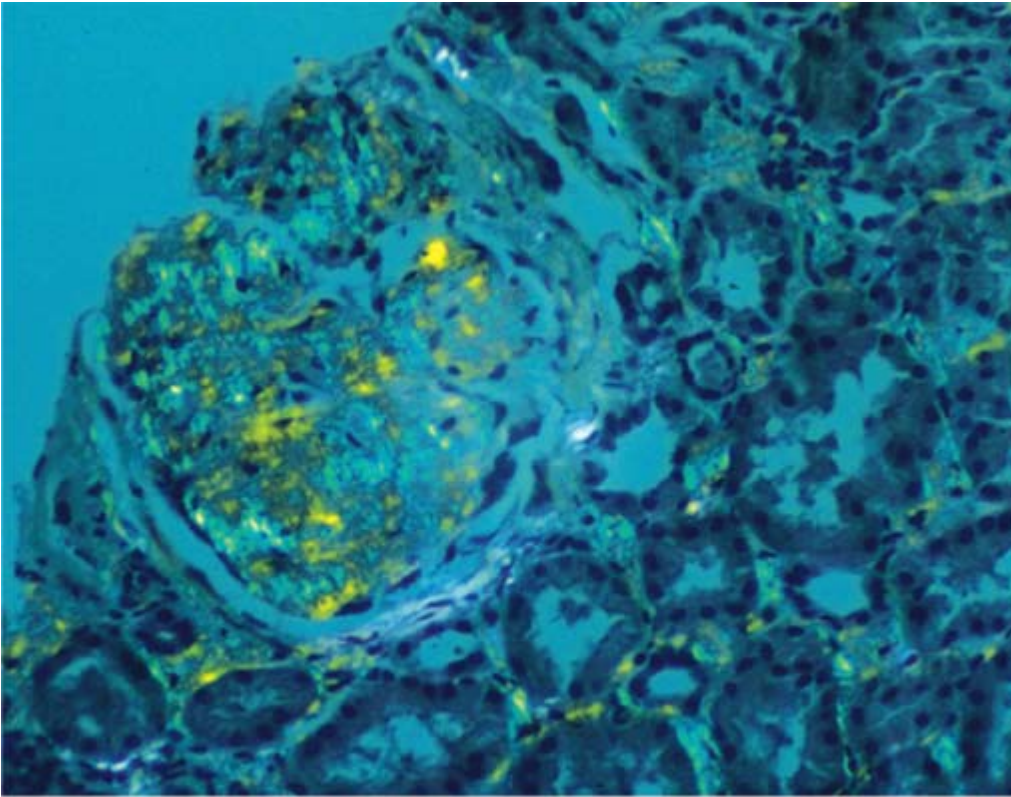
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

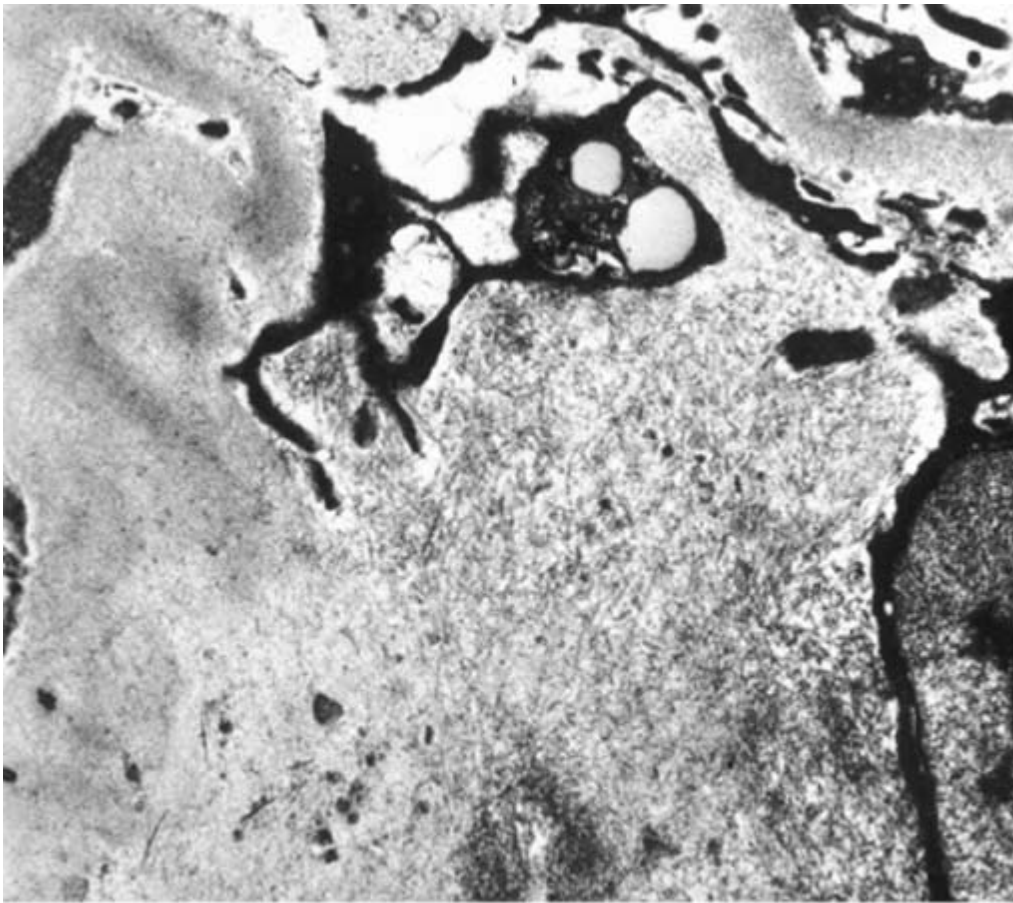
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-18

Amyloidosis. Amyloidosis shows amorphous, acellular expansion of the mesangium, with material often also infiltrating glomerular basement membranes, vessels, and the interstitium, with apple-green birefringence by polarized Congo red stain (**A**). The deposits are composed of randomly organized 9–11 nm fibrils by electron microscopy (**B**). (*ABF/Vanderbilt Collection.*)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

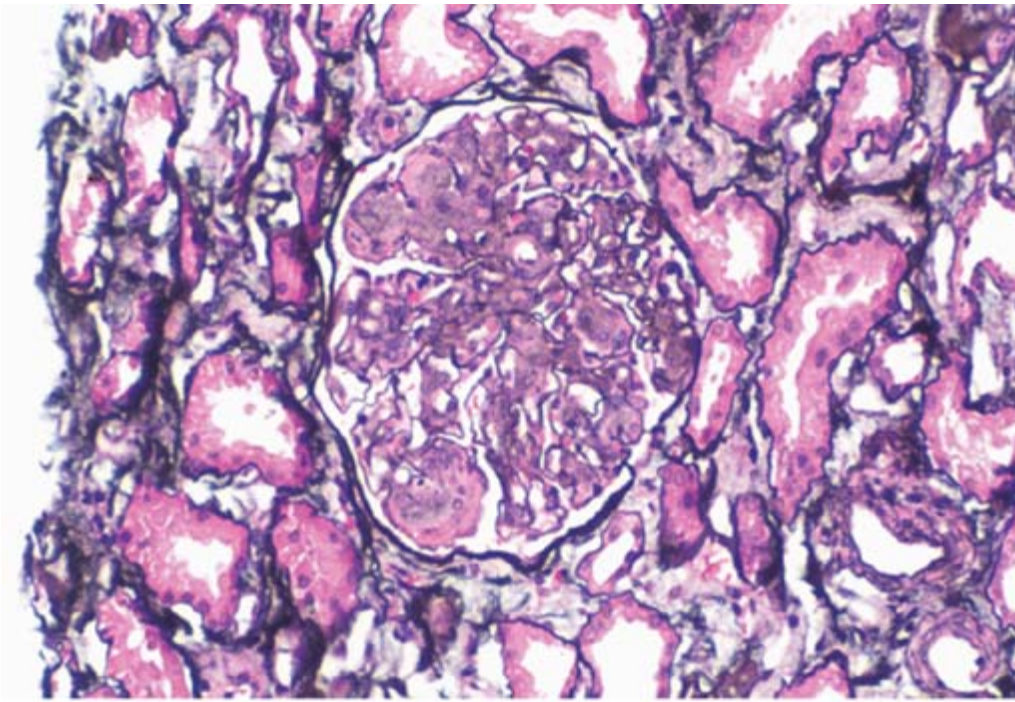


B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

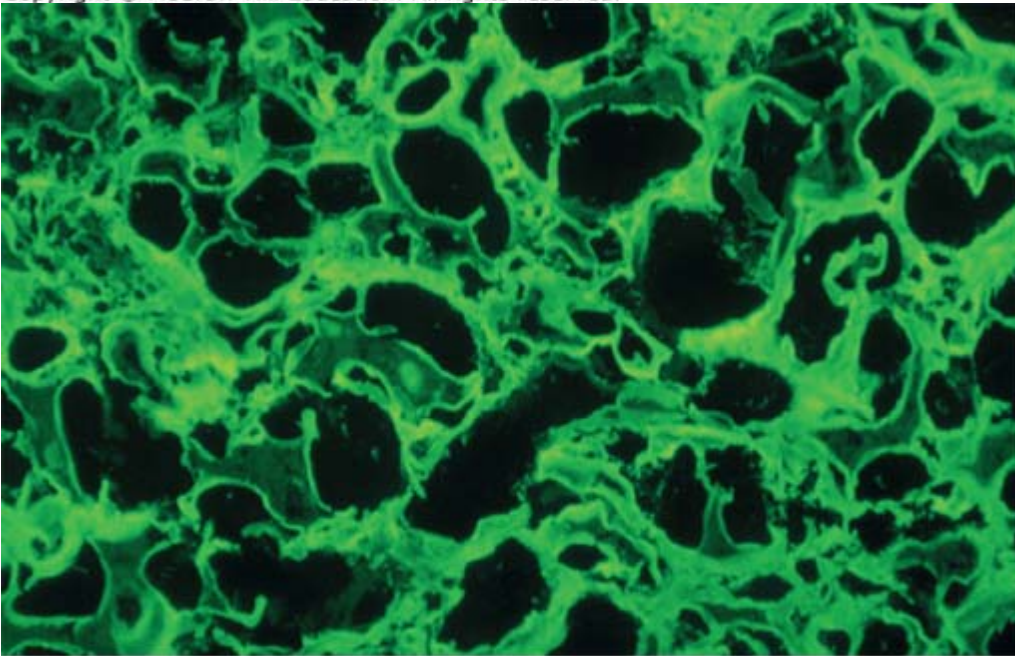
FIGURE A3-19

Light chain deposition disease. There is mesangial expansion, often nodular by light microscopy (**A**), with immunofluorescence showing monoclonal light chain staining, more commonly with kappa than lambda light chain, of tubules (**B**) and glomerular tufts. By electron microscopy (**C**), the deposits show an amorphous granular appearance and line the inside of the glomerular basement membrane (*arrows*) and are also found along the tubular basement membranes. (*ABF/Vanderbilt Collection.*)



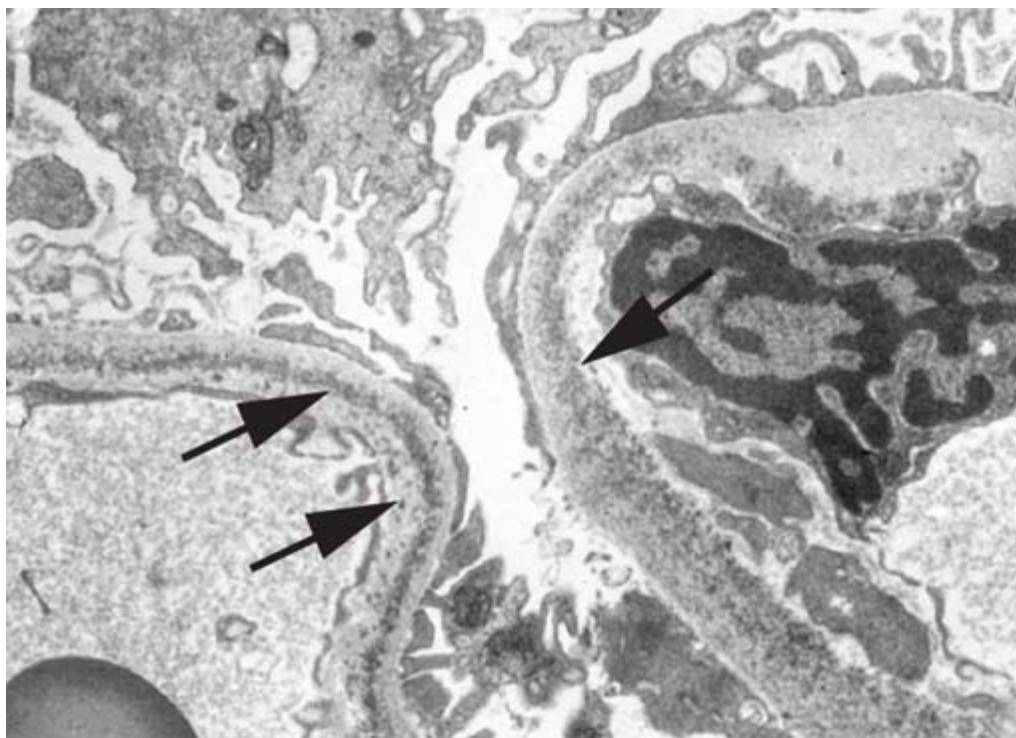
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

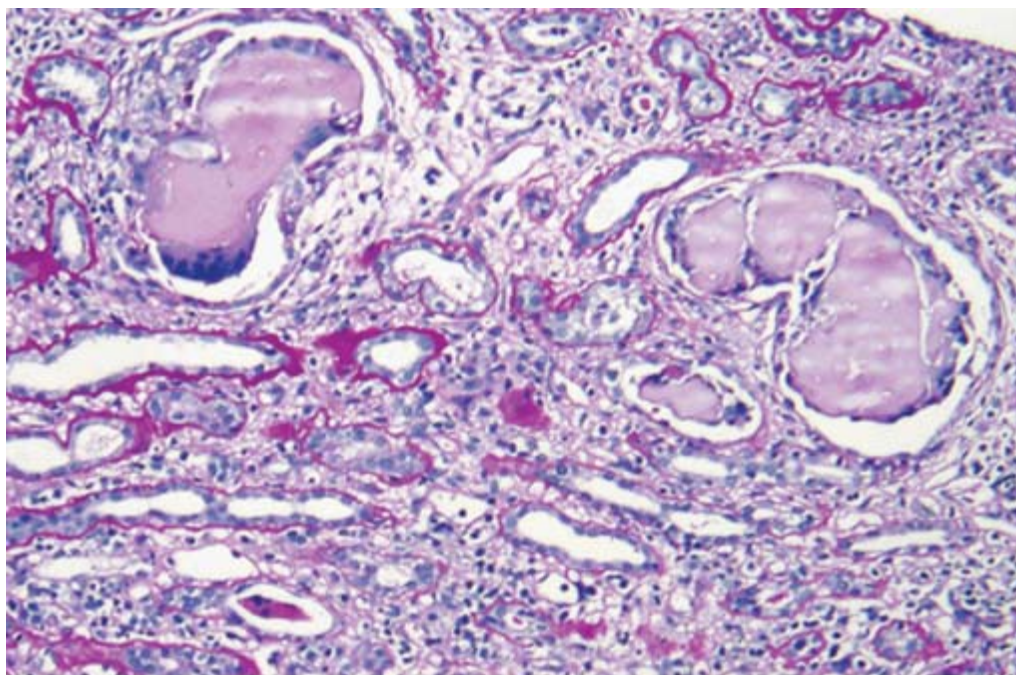
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

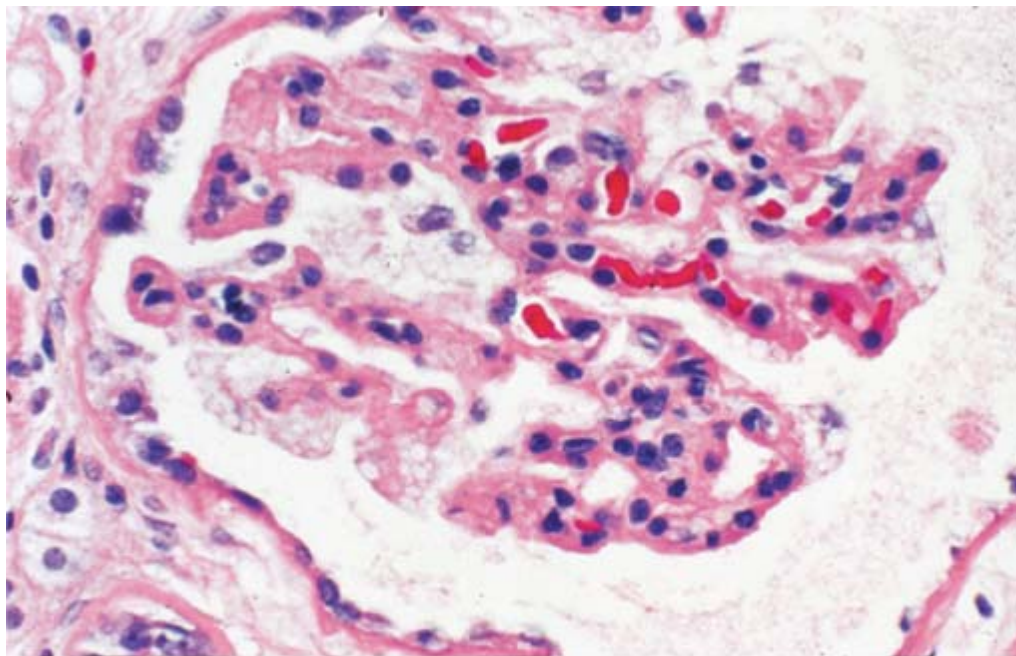
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-20

Light chain cast nephropathy (myeloma kidney). Monoclonal light chains precipitate in tubules and result in a syncytial giant cell reaction surrounding the casts and a surrounding chronic interstitial nephritis with tubulointerstitial fibrosis. (*ABF/Vanderbilt Collection.*)



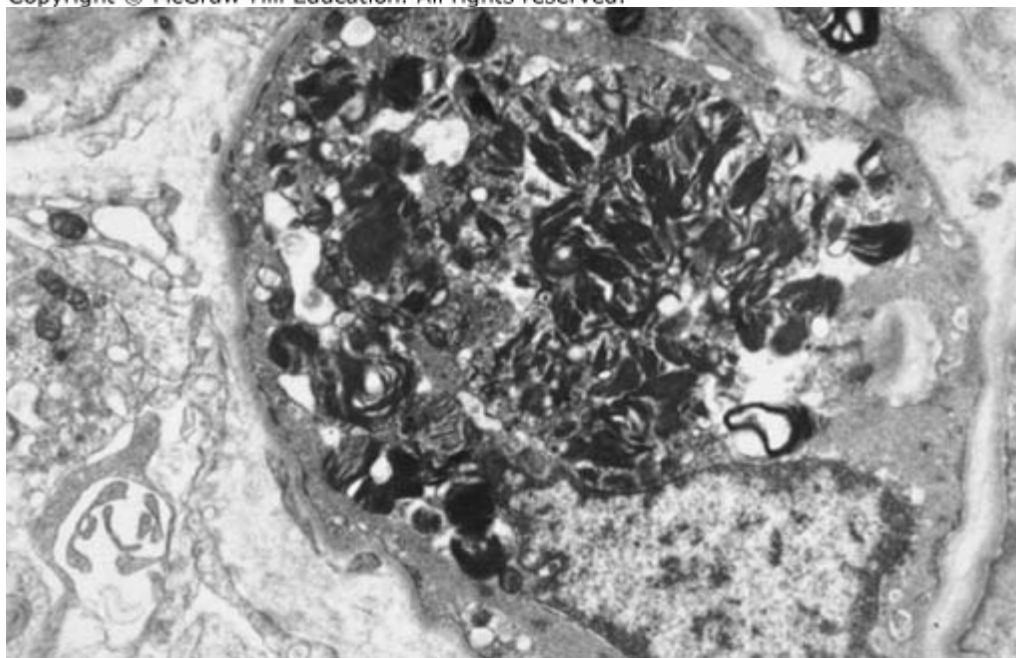
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-21

Fabry's disease. Due to deficiency of α -galactosidase, there is abnormal accumulation of glycolipids, resulting in foamy podocytes by light microscopy (**A**). These deposits can be directly visualized by electron microscopy (**B**), where the glycosphingolipid appears as whorled so-called myeloid bodies, particularly in the podocytes. (ABF/Vanderbilt Collection.)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

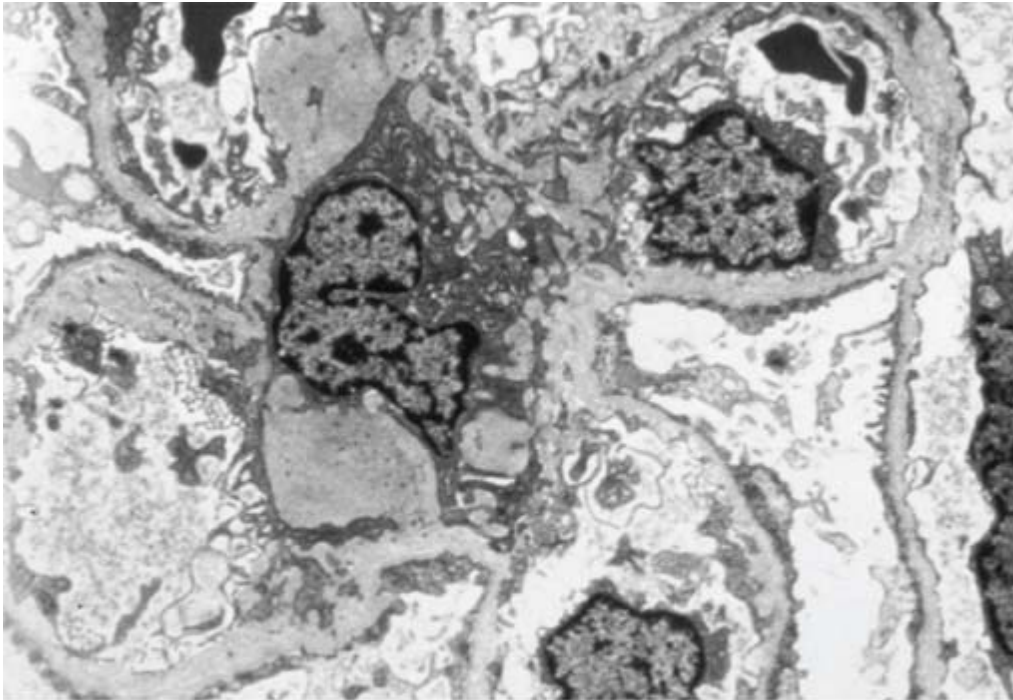


B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

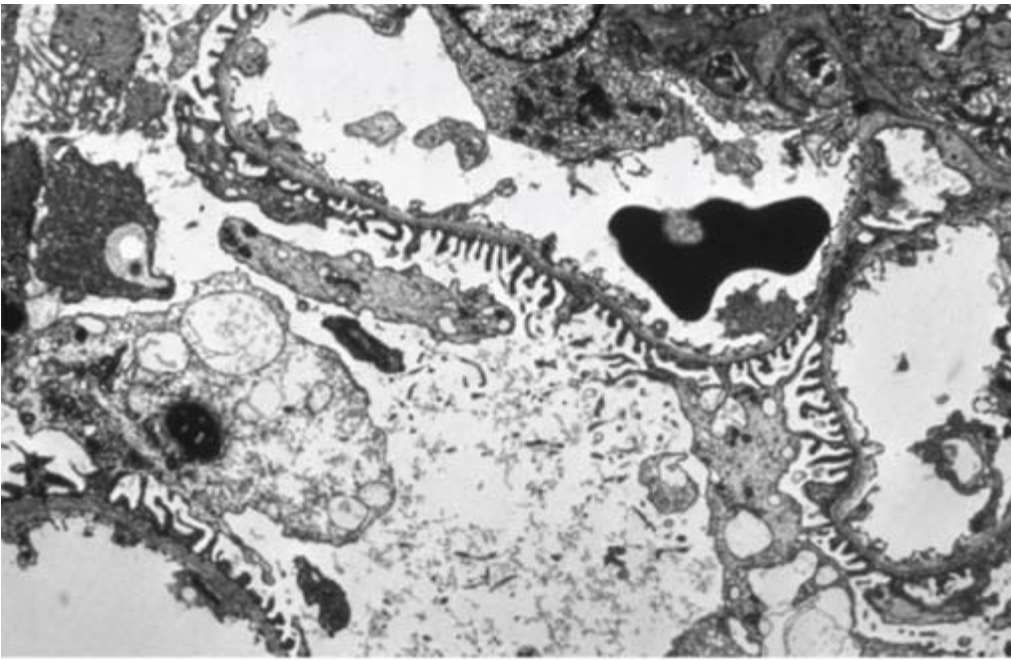
FIGURE A3-22

Alport's syndrome and thin glomerular basement membrane lesion. In Alport's syndrome, there is irregular thinning alternating with thickened so-called basket-weaving abnormal organization of the glomerular basement membrane (**A**). In benign familial hematuria (often due to a carrier state of autosomal recessive Alport), or in early cases of Alport's syndrome or female carriers, only extensive thinning of the glomerular basement membrane is seen by electron microscopy (**B**). (*ABF/Vanderbilt Collection.*)



A

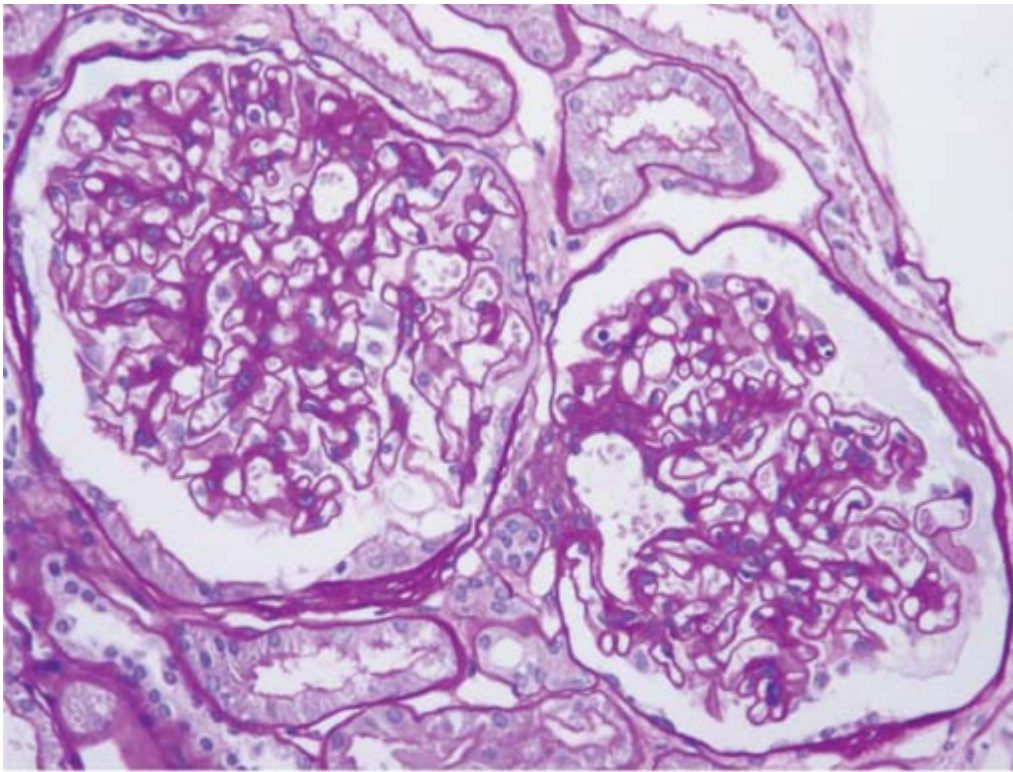
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

**B**

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
 Copyright © McGraw-Hill Education. All rights reserved.

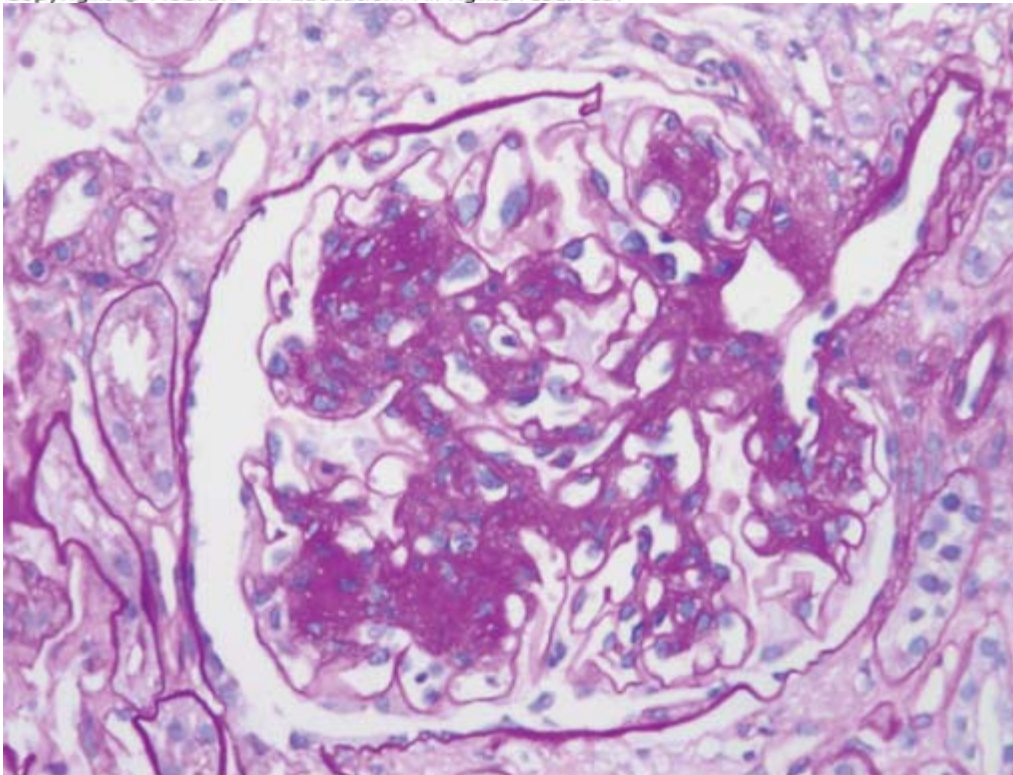
FIGURE A3-23

Diabetic nephropathy. In the earliest stage of diabetic nephropathy, only mild mesangial increase and prominent glomerular basement membranes (confirmed to be thickened by electron microscopy) are present (**A**). In slightly more advanced stages, more marked mesangial expansion with early nodule formation develops, with evident arteriolar hyaline (**B**). In established diabetic nephropathy, there is nodular mesangial expansion, so-called Kimmelstiel-Wilson nodules, with increased mesangial matrix and cellularity, microaneurysm formation in the glomerulus on the left, and prominent glomerular basement membranes without evidence of immune deposits and with hyalinosis of both afferent and efferent arterioles (**C**).
 (ABF/Vanderbilt Collection.)



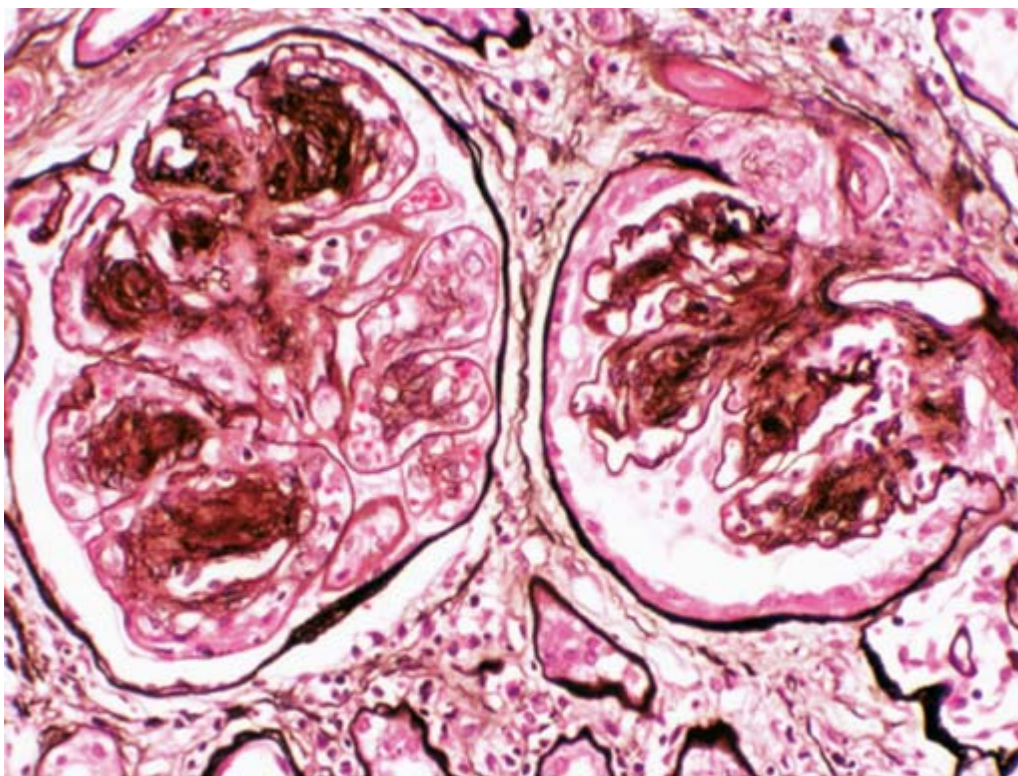
A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

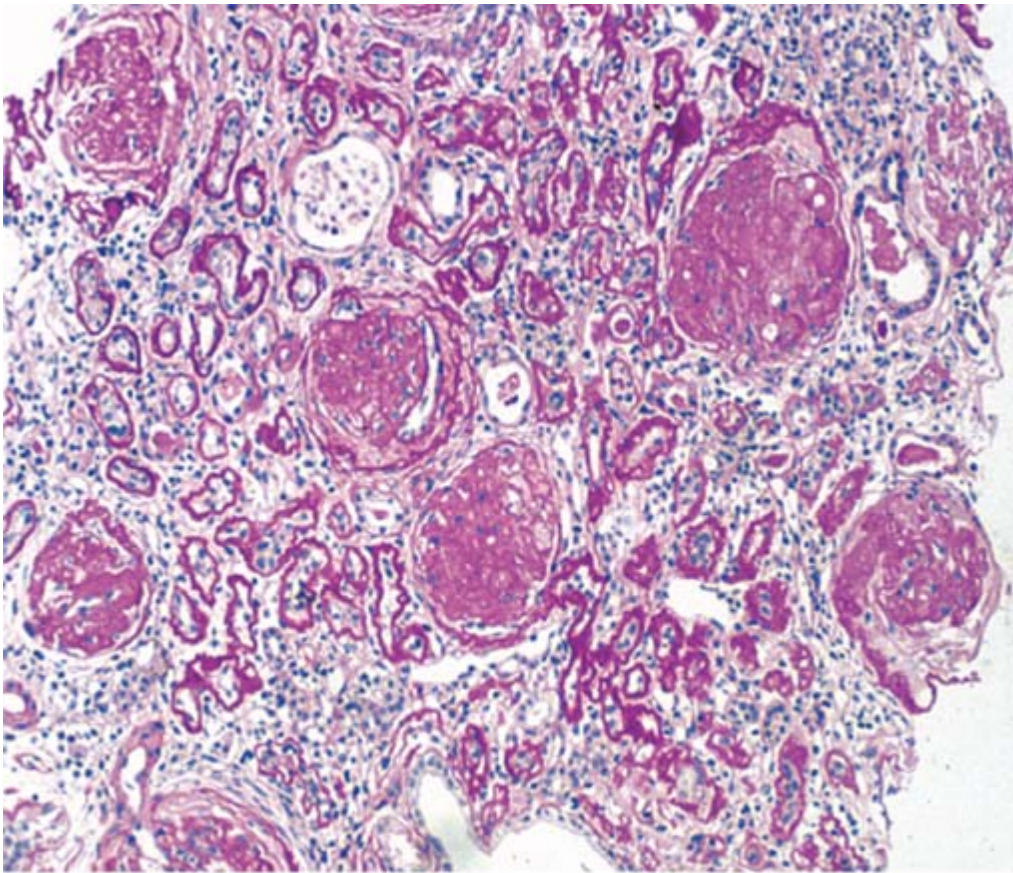
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



C

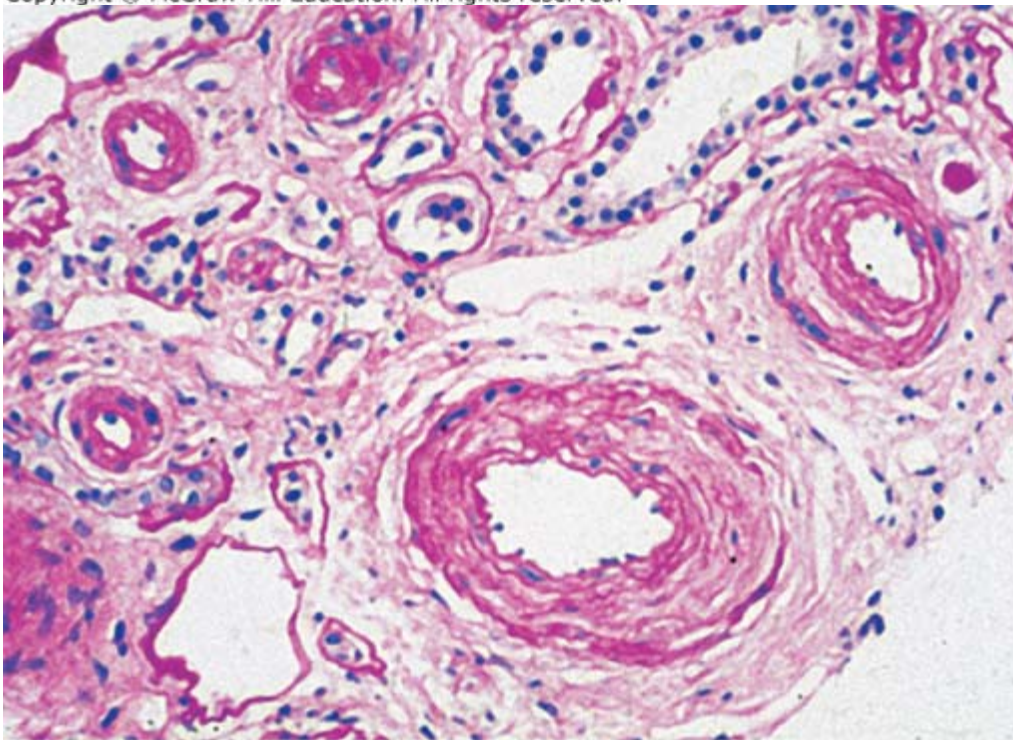
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-24

Arterionephrosclerosis. Hypertension-associated injury often manifests extensive global sclerosis of glomeruli, with accompanying and proportional tubulointerstitial fibrosis and pericapsular fibrosis, and there may be segmental glomerulosclerosis (**A**). The vessels show disproportionately severe changes of intimal fibrosis, medial hypertrophy, and arteriolar hyaline deposits (**B**). (*ABF/Vanderbilt Collection.*)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

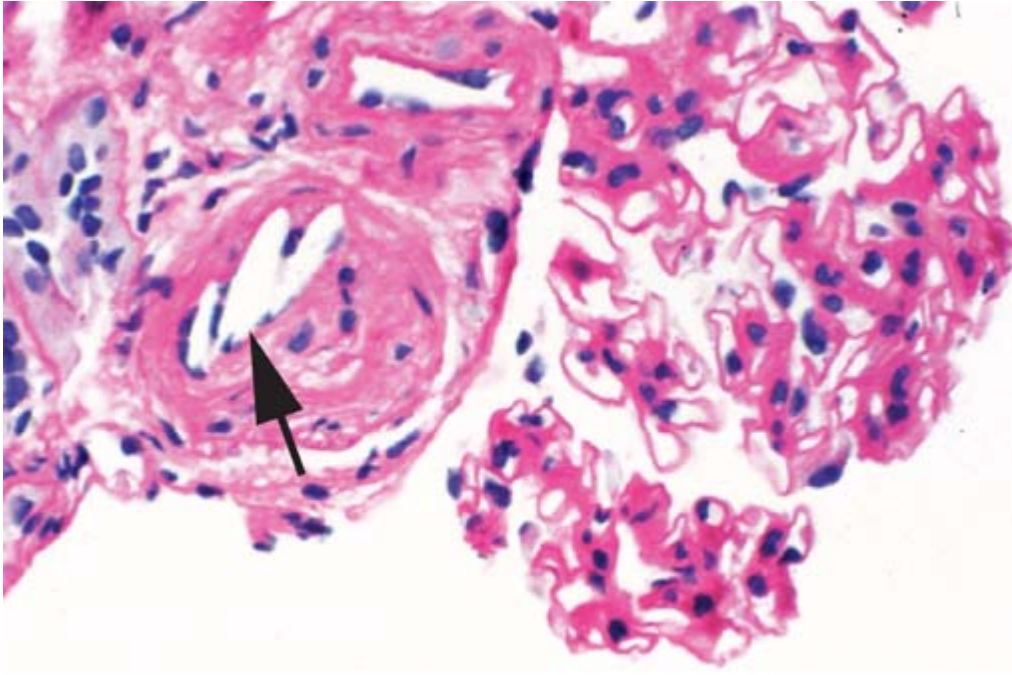


B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-25

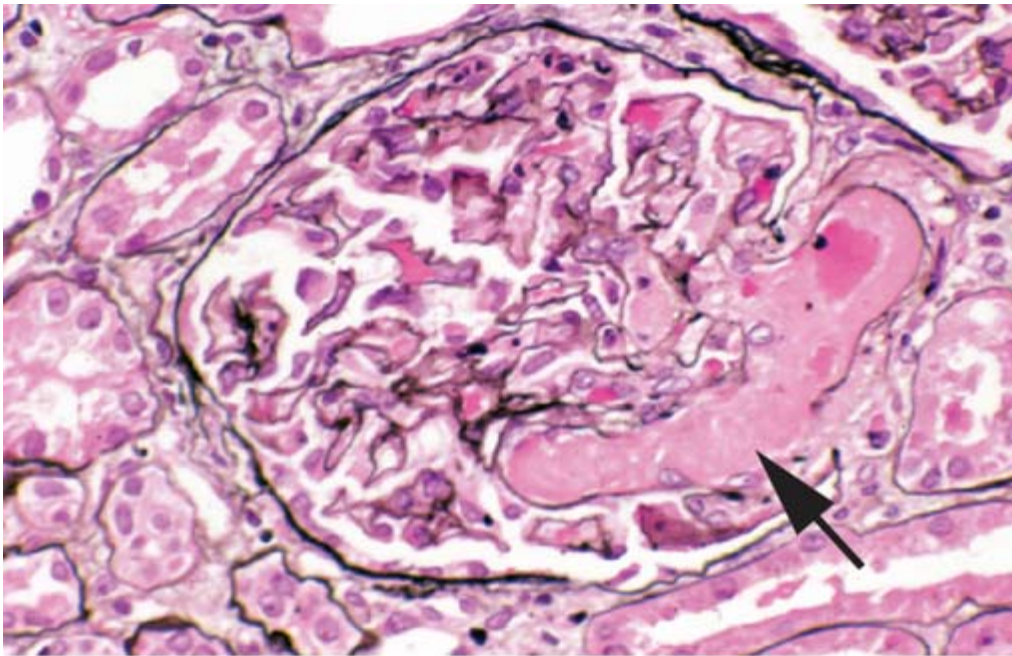
Cholesterol emboli. Cholesterol emboli cause cleft-like spaces (*arrow*) where the lipid has been extracted during processing, with smooth outer contours and surrounding fibrotic and mononuclear cell reaction in these arterioles. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

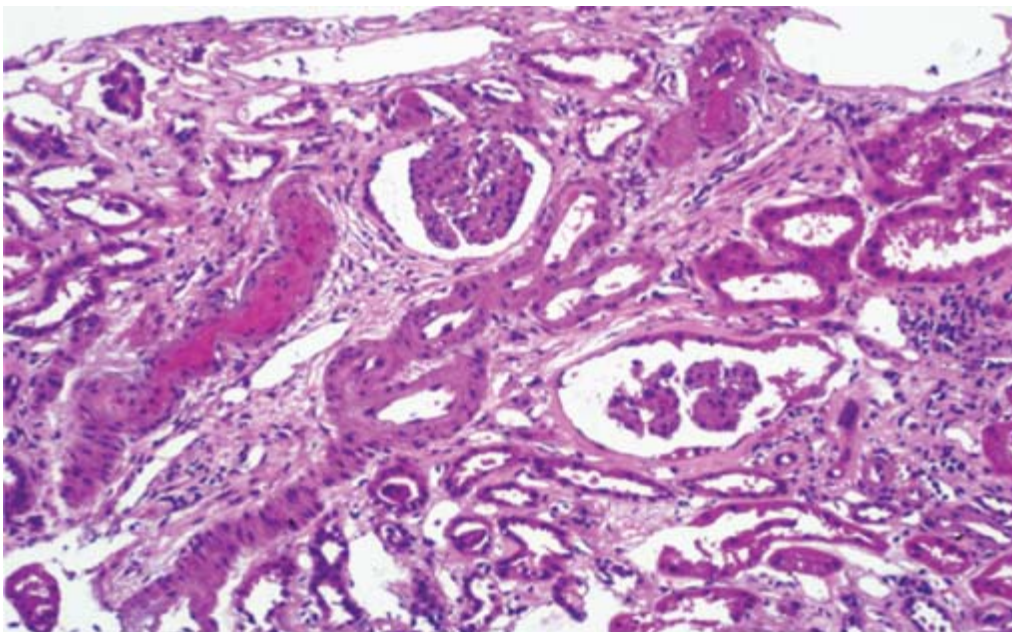
FIGURE A3-26

Thrombotic microangiopathy. This is the classic lesion in hemolytic uremic syndrome, with characteristic intraglomerular fibrin thrombi, with a chunky pink appearance (*arrow*). The remaining portion of the capillary tuft shows corrugation of the glomerular basement membrane due to ischemia. (*ABF/Vanderbilt Collection.*)



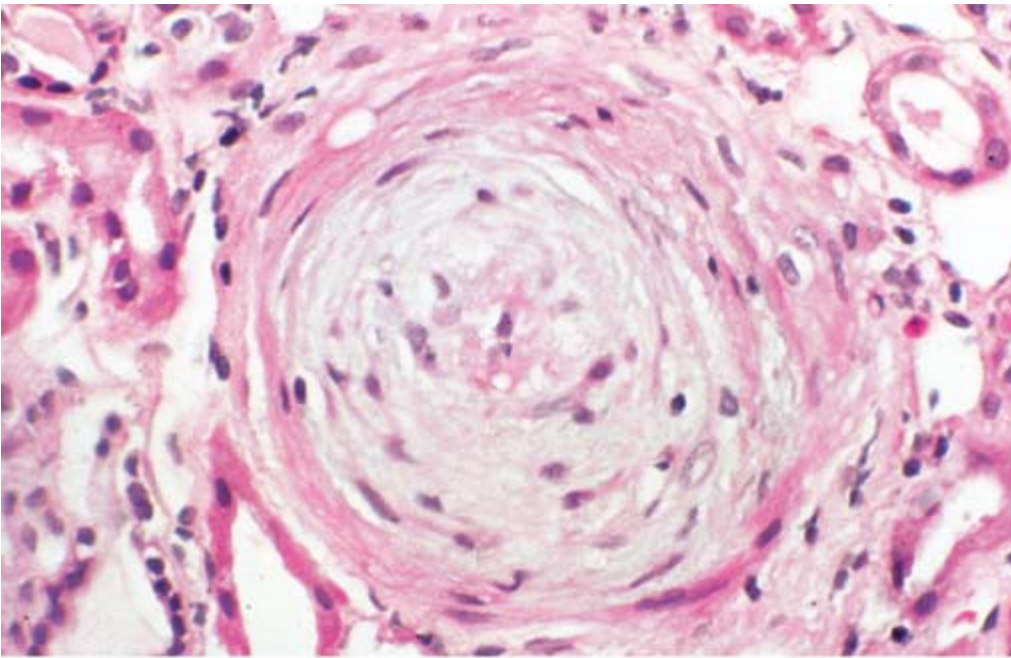
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-27

Systemic sclerosis. Acutely, there is fibrinoid necrosis of interlobular and larger vessels, with intervening normal vessels and ischemic change in the glomeruli, manifest by corrugation of the glomerular basement membranes (**A**). Chronically, this injury leads to intimal proliferation, the so-called onion-skinning appearance (**B**). (*ABF/Vanderbilt Collection.*)



A

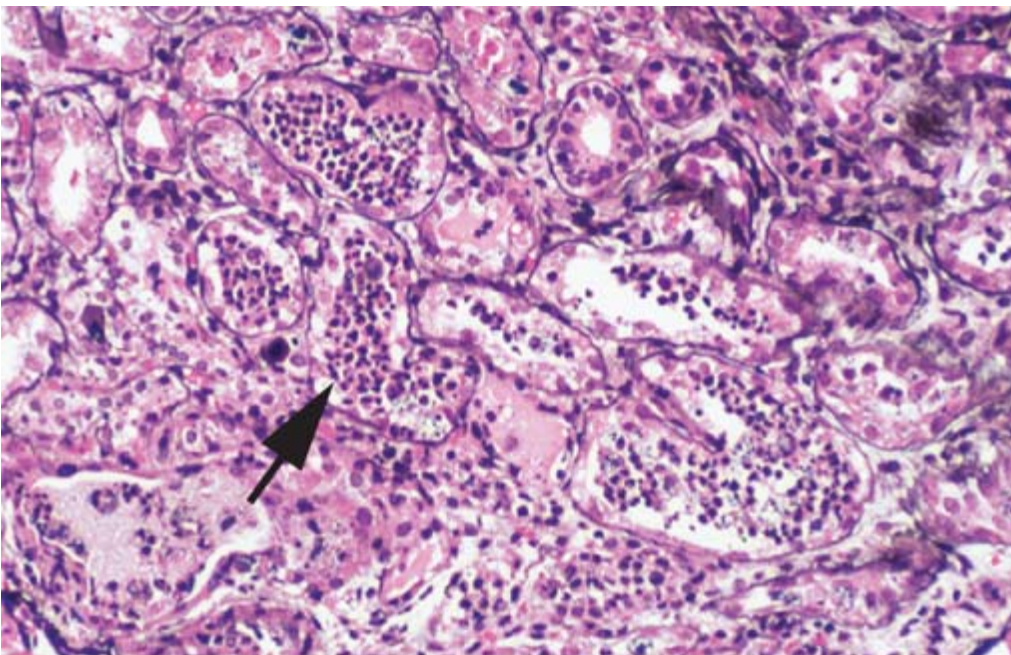
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

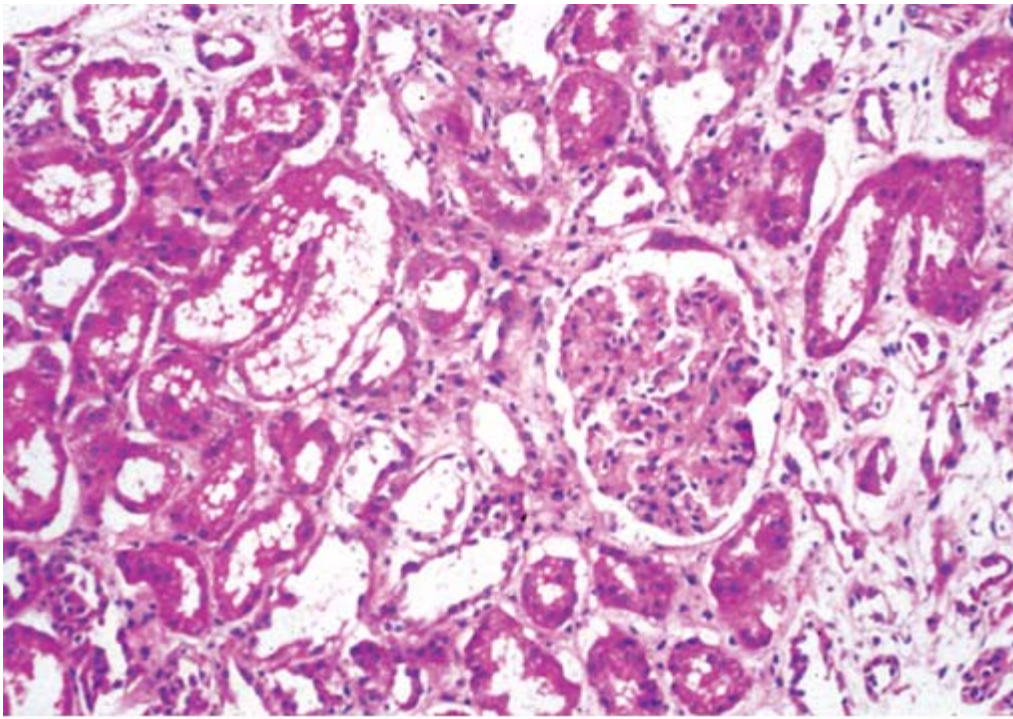
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-28

Acute pyelonephritis. There are characteristic intratubular plugs and casts of polymorphonuclear neutrophils (PMNs) (*arrow*) with inflammation extending into the surrounding interstitium and accompanying tubular injury. (*ABF/Vanderbilt Collection.*)



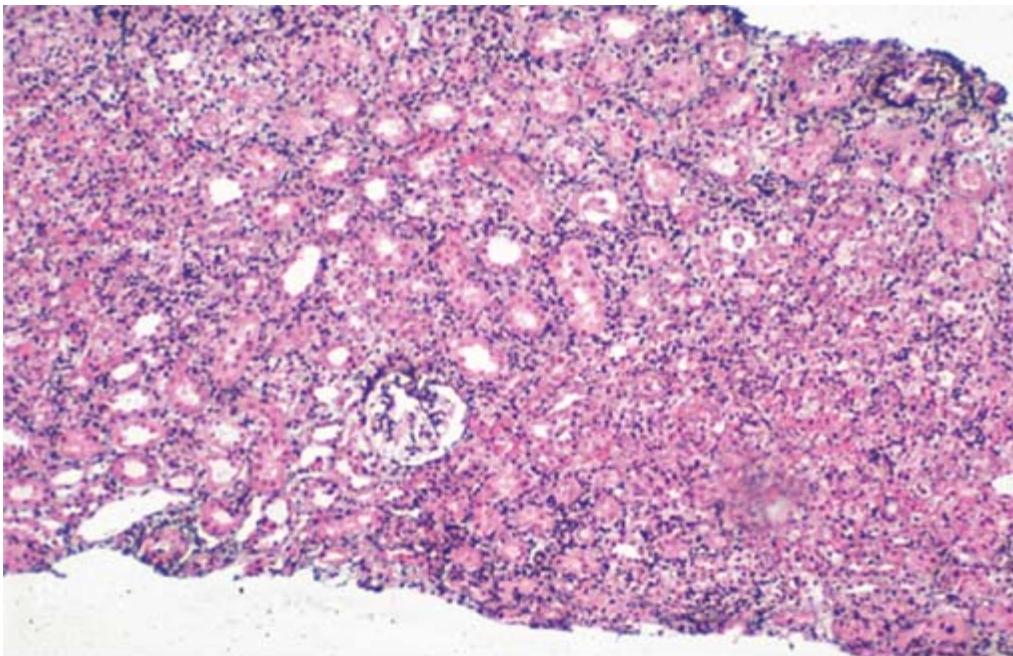
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-29

Acute tubular injury. There is extensive flattening of the tubular epithelium and loss of the brush border, with mild interstitial edema, characteristic of acute tubular injury due to ischemia. (*ABF/Vanderbilt Collection.*)



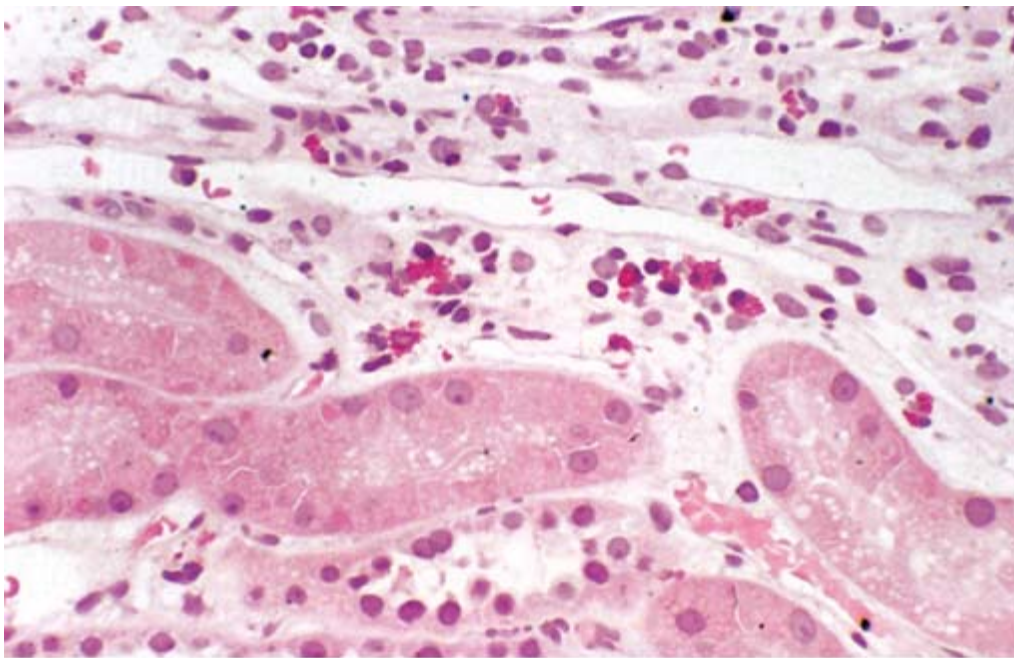
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-30

Acute interstitial nephritis. There is extensive interstitial lymphoplasmocytic infiltrate with mild edema and associated tubular injury (**A**), which is frequently associated with interstitial eosinophils (**B**) when caused by a drug hypersensitivity reaction. (*ABF/Vanderbilt Collection.*)



A

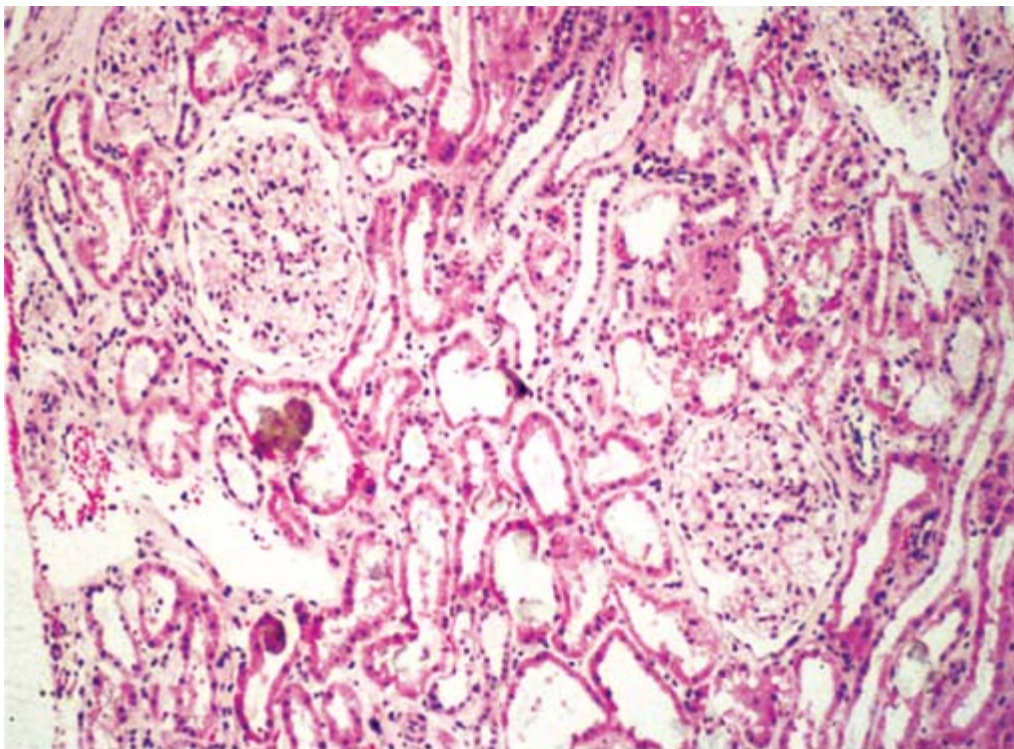
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.



B

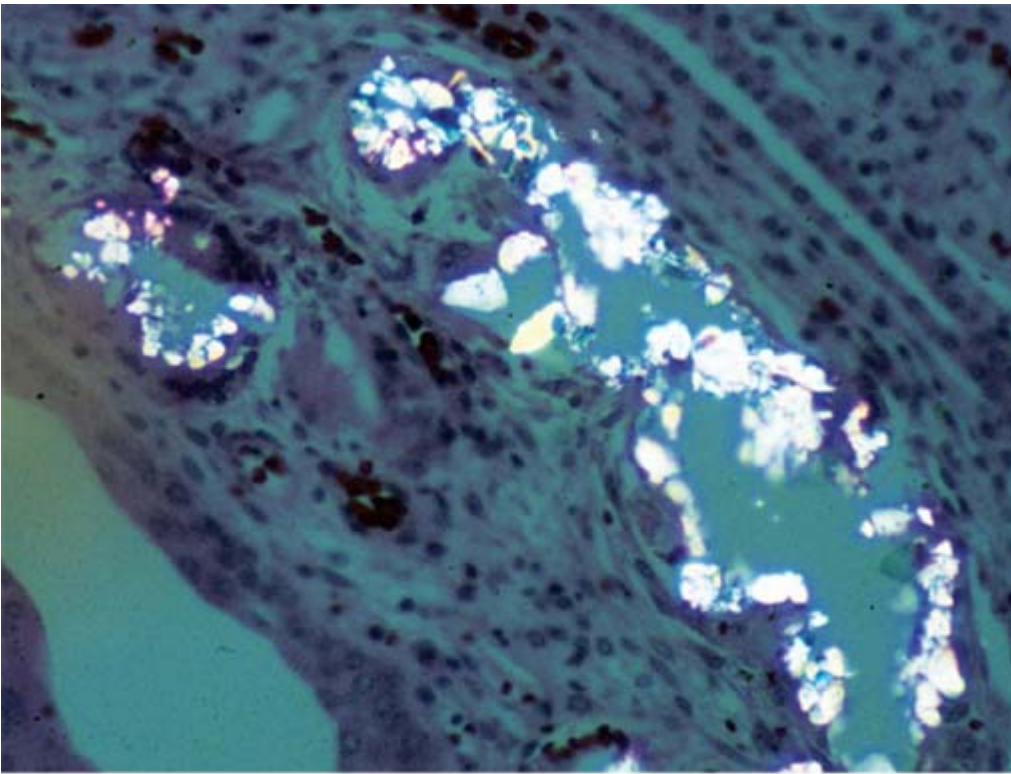
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-31

Oxalosis. Calcium oxalate crystals have caused extensive tubular injury, with flattening and regeneration of tubular epithelium (**A**). Crystals are well visualized as sheaves when viewed under polarized light (**B**).
(ABF/Vanderbilt Collection.)



A

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

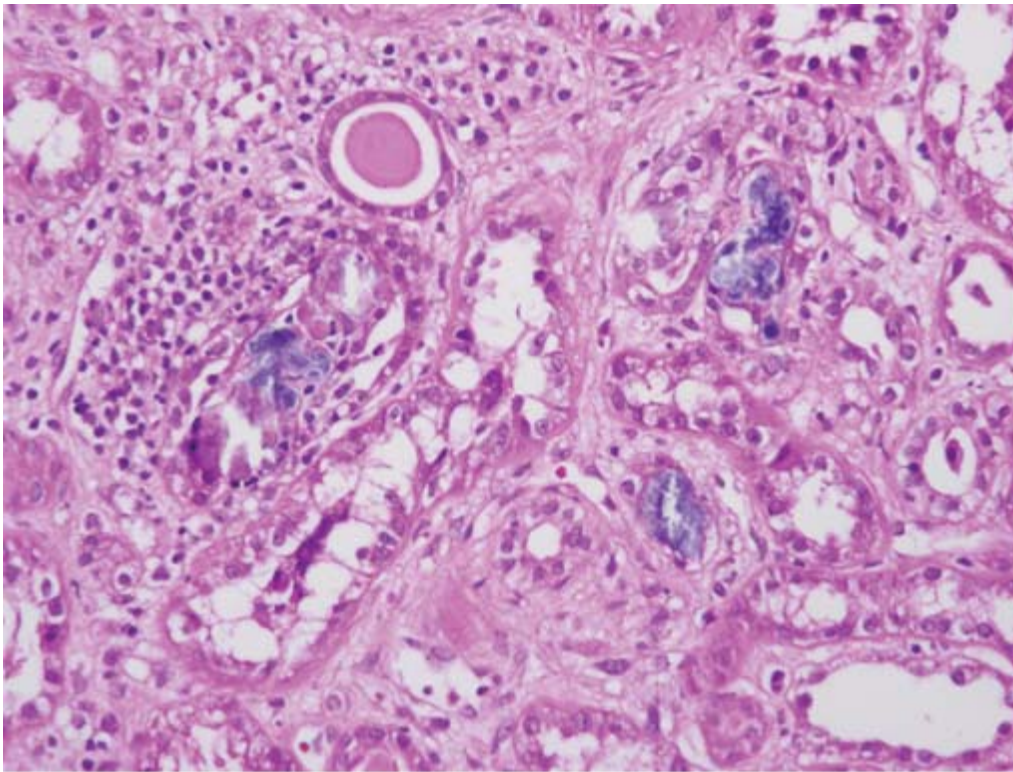


B

Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

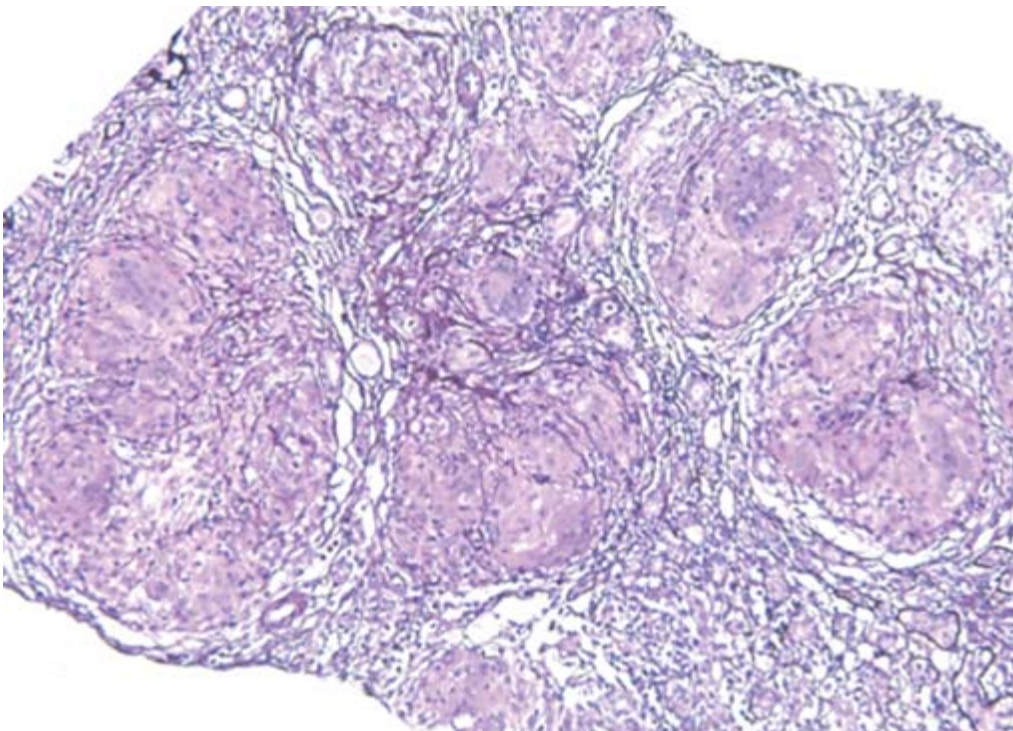
FIGURE A3-32

Acute phosphate nephropathy. There is extensive acute tubular injury with intratubular nonpolarizable calcium phosphate crystals. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-33

Sarcoidosis. There is chronic interstitial nephritis with numerous, confluent, nonnecrotizing granulomas. The glomeruli are unremarkable, but there is moderate tubular atrophy and interstitial fibrosis. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-34

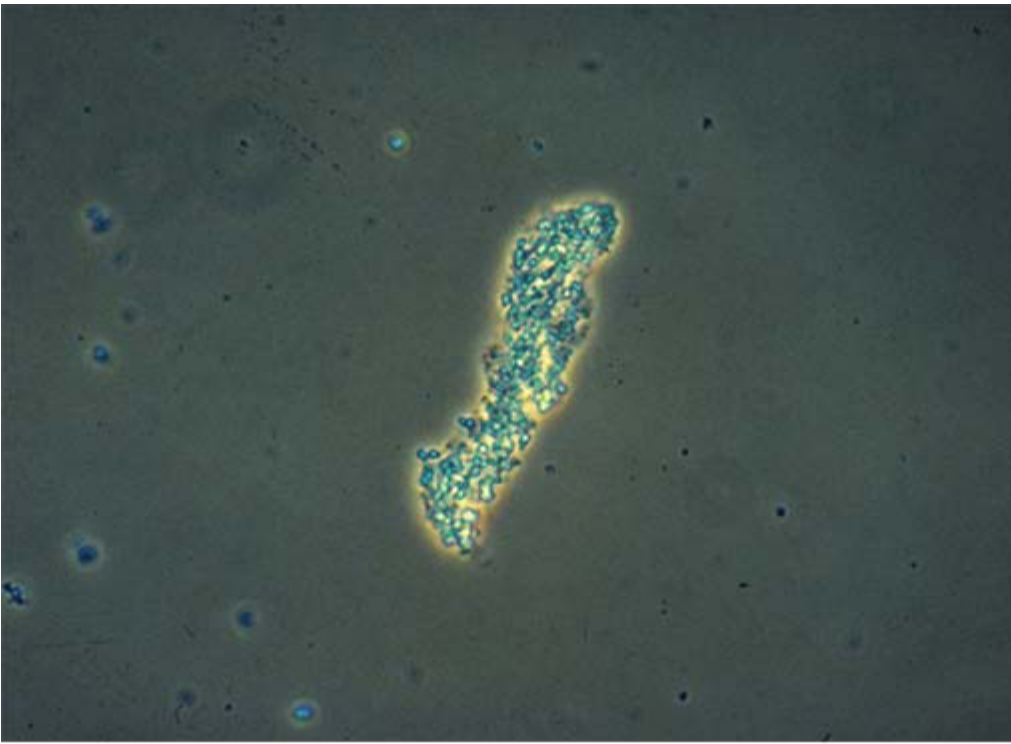
Hyaline cast. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

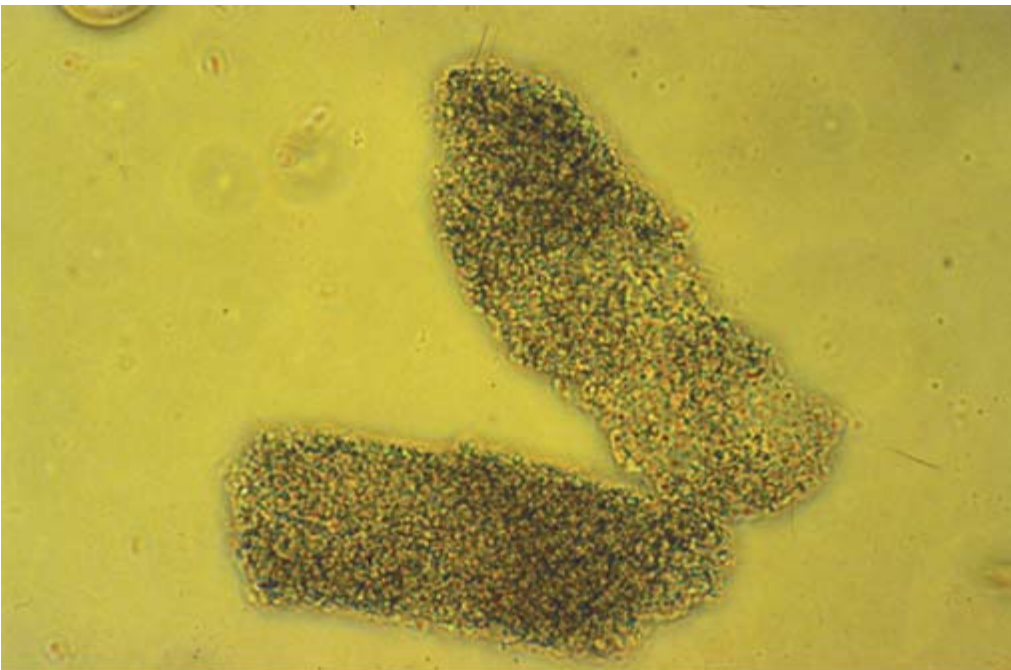
FIGURE A3-35

Coarse granular cast. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-36

Fine granular casts. (*ABF/Vanderbilt Collection.*)



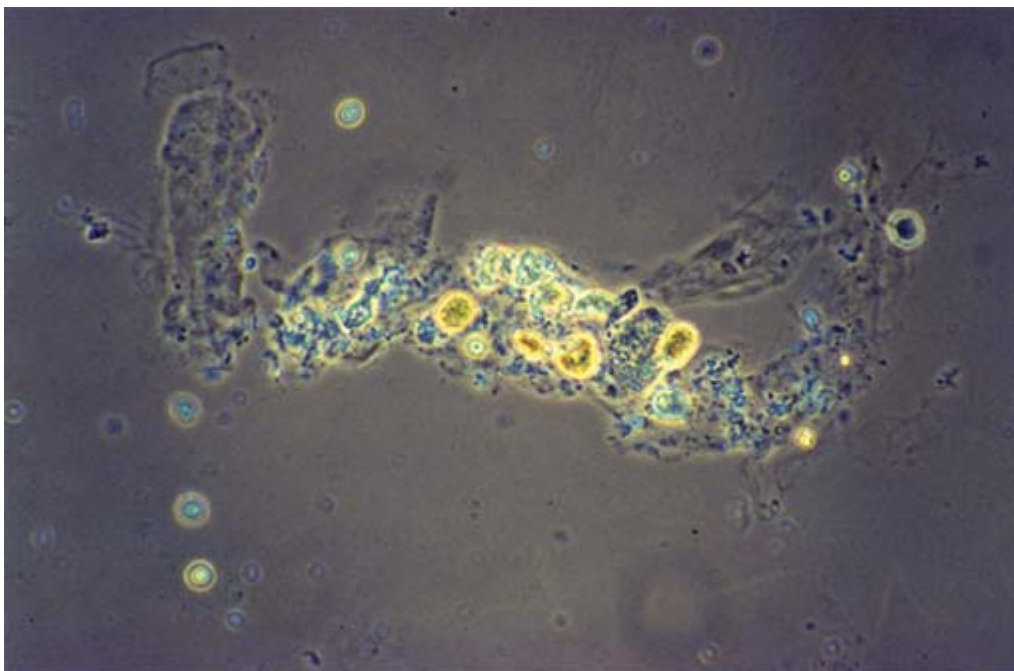
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-37

Red blood cell cast. (*ABF/Vanderbilt Collection.*)



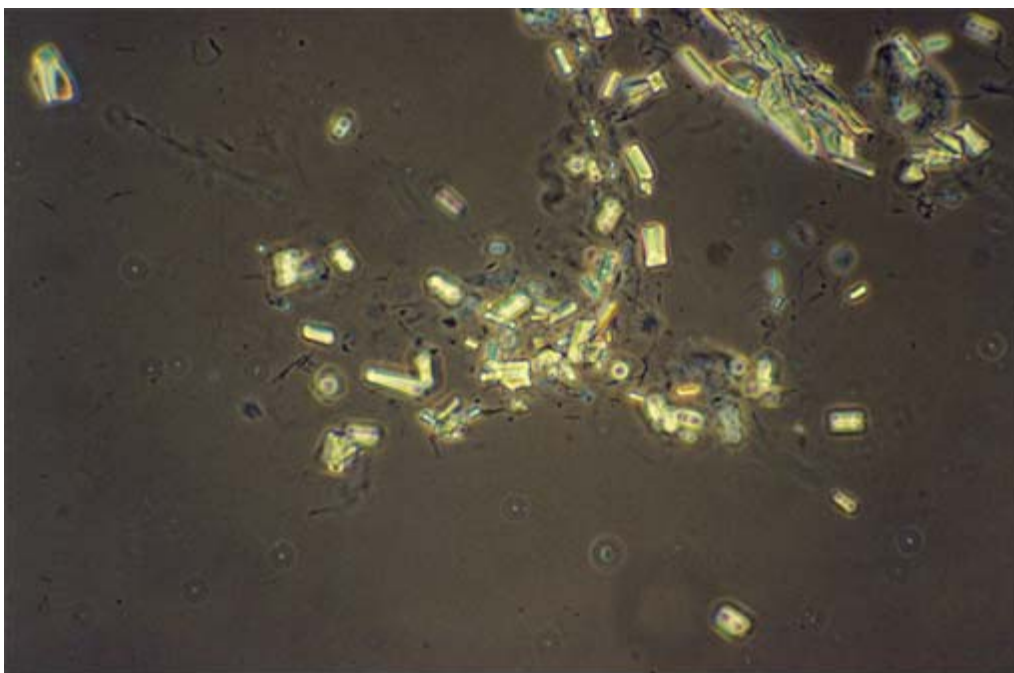
Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-38

White blood cell cast. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.
FIGURE A3-39

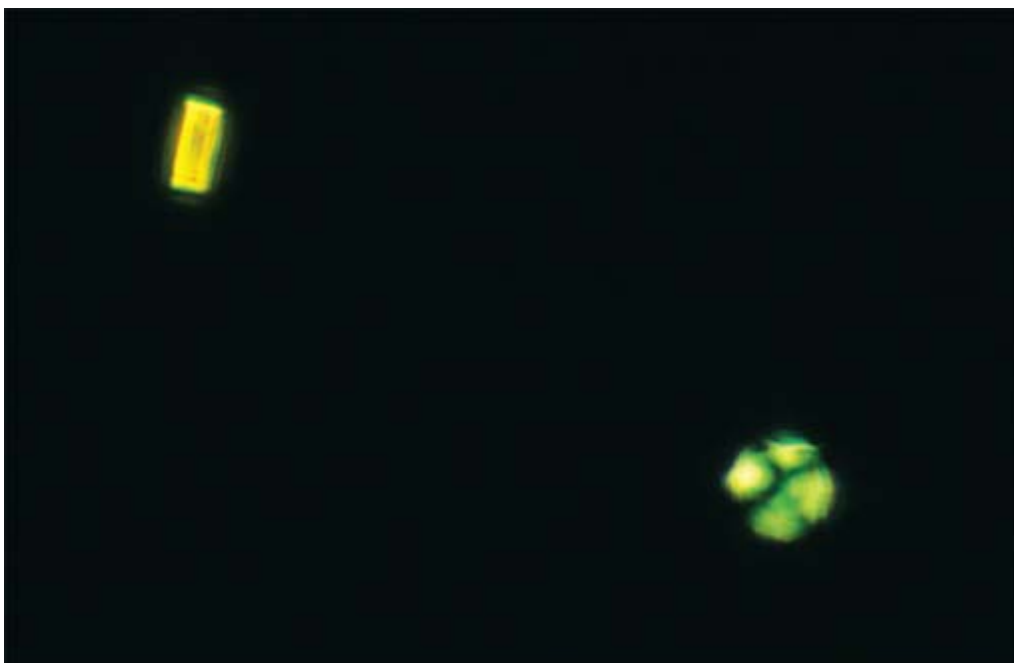
Triple phosphate crystals. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-40

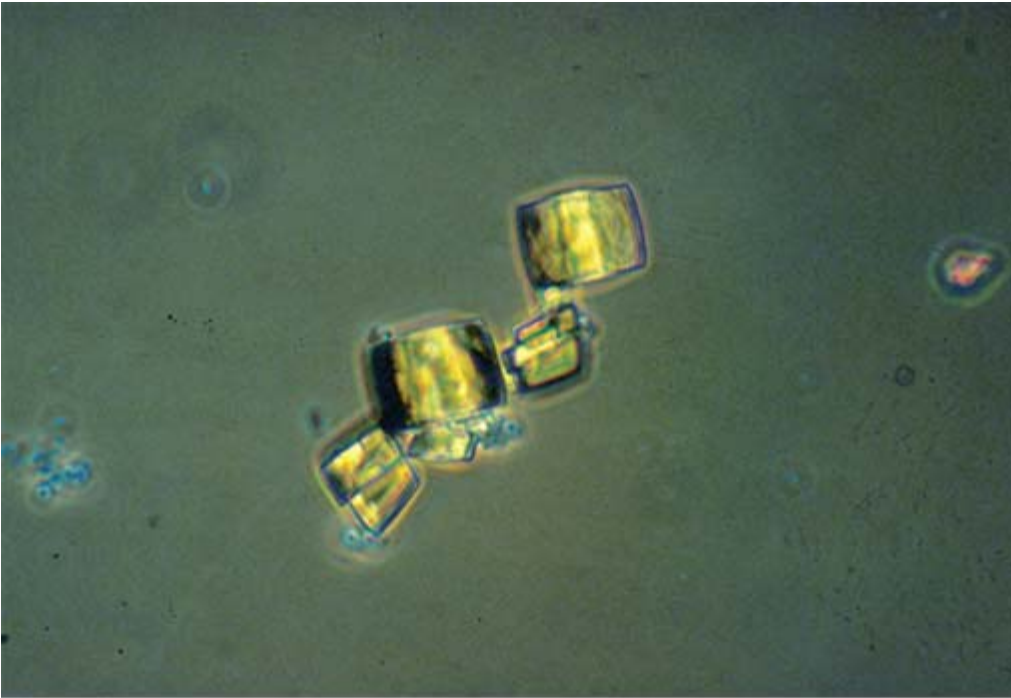
“Maltese cross” formation in an oval fat body. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

FIGURE A3-41

Uric acid crystals. (*ABF/Vanderbilt Collection.*)



Source: J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition
Copyright © McGraw-Hill Education. All rights reserved.

[McGraw Hill](#)

Copyright © McGraw-Hill Global Education Holdings, LLC.

All rights reserved.

Your IP address is **178.250.250.21**

Access Provided by: National Medical Library

[Silverchair](#)